

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:56:11 ; Search time 112.99 Seconds
 (without alignments)
 2709.566 Million cell updates/sec

Title: US-09-361-619-11
 Perfect score: 11694
 Sequence: 1 MNHLYKVFKNKATGTFMAVA.....NGSADTOGHVGAAGVGFHF 2314

Scoring table:
 BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL_16.*
 1: sp.archaea.*
 2: sp.bacteria.*
 3: sp.fungi.*
 4: sp.human.*
 5: sp.invertebrate.*
 6: sp.mammal.*
 7: sp.mnc.*
 8: sp.organelle.*
 9: sp.phage.*
 10: sp.plant.*
 11: sp.rodent.*
 12: sp.unclassified.*
 13: sp.vertibrate.*
 14: sp.virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	1222.5	10.5	2059	2 Q9PD50	Q9pd50 xylella fas
2	1177.5	10.1	2353	2 P71401	P71401 haemophilus
3	1037	8.9	2712	2 Q9F3X5	Q9f3x5 pasteurella
4	743.5	6.4	1190	2 Q9PC04	Q9pc04 xylella fas
5	697.5	6.0	3705	2 Q9F2D8	Q9f2d8 versinia pe
6	661.5	5.7	1107	2 Q9F2D8	Q9f2d8 salmonella
7	659.5	5.6	1098	2 Q48152	Q48152 haemophilus
8	653.5	5.6	1299	2 Q9F3X6	Q9f3x6 pasteurella
9	588	5.0	2586	5 Q9VTK8	Q9vtk8 drosophila
10	578	4.9	2478	2 Q9RL69	Q9rl69 staphylococ
11	578	4.9	2478	2 Q9LCH2	Q9lch2 staphylococ
12	571.5	4.9	2340	2 Q9ZD91	Q9zd91 rickettsia
13	565.5	4.8	2021	2 Q52657	Q52657 rickettsia
14	560	4.8	3029	2 Q55582	Q55582 synecocyst
15	558.5	4.8	2747	2 Q91800	Q91800 aeromonas s
16	552.5	4.7	4919	2 Q9ZHL0	Q9zhl0 haemophilus
17	546	4.7	2514	2 Q9YJ30	Q9yj30 neisseria m
18	536.5	4.6	5627	2 Q91120	Q91120 pseudomonas
19	534	4.6	1557	2 Q9RNI2	Q9rni2 haemophilus

ALIGNMENTS

RESULT 1

ID	Q9PD50	PRELIMINARY;	PRT;	2059 AA.
AC	Q9PD50;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)			
DE	SURFACE PROTEIN.			
GN	XF1529.			
OS	Xylella fastidiosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;			
OC	Xylella.			
OX	NCBI_TaxID=2371;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=9A5C;			
RX	MEDLINE=20365717; PubMed=10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,			
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,			
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Cartaro D.M., Carter H.,			
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,			
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfy H.,			
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,			
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,			
RA	Machado M.A., Madeira A.M.B.N., Madeira E.M.F., Marino C.L.,			
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,			
RA	Mench C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,			
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,			
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Sawasaki H.E.,			
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,			
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,			
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,			

Q91950 pseudomonas
 O18758 sus scrofa
 Q9xc47 rickettsia
 Q9i2m3 pseudomonas
 Q9kb5 rickettsia
 Q9xcj4 salmonella
 Q9k0t0 neisseria m
 Q9xcq3 salmonella
 Q9zhl3 haemophilus
 Q9kk99 rickettsia
 Q9kb1 rickettsia
 Q48031 haemophilus
 Q9jms5 escherichia
 Q9kka5 rickettsia
 Q48028 haemophilus
 Q9kka9 rickettsia
 Q9f0p9 rickettsia
 Q9kka7 rickettsia
 Q9kka4 rickettsia
 Q9kka2 rickettsia
 Q9kka1 rickettsia
 Q9kka3 rickettsia
 Q9fda0 xanthomonas
 Q9kba0 rickettsia
 P94750 escherichia

20 527 4.5 2147 2 Q91950
 21 520.5 4.5 13288 6 O18758
 22 516 4.4 2106 2 Q9xc47
 23 511 4.4 2468 2 Q9i2m3
 24 509 4.4 1620 2 Q9kb5
 25 506.5 4.3 2035 2 Q9xcj4
 26 504.5 4.3 2703 2 Q9k0t0
 27 504 4.3 1963 2 Q9xcq3
 28 504 4.3 4152 2 Q9zhl3
 29 492.5 4.2 1604 2 Q9kk99
 30 492.5 4.2 1618 2 Q9kb1
 31 483 4.1 1536 2 Q48031
 32 482 4.1 1758 2 Q9jms5
 33 475.5 4.1 1616 2 Q9kka5
 34 472 4.0 1477 2 Q48028
 35 470 4.0 1616 2 Q9kka9
 36 469 4.0 1615 2 Q9f0p9
 37 469 4.0 1616 2 Q9kka7
 38 468.5 4.0 1618 2 Q9kka4
 39 467.5 4.0 1617 2 Q9kka2
 40 465 4.0 1616 2 Q9kka1
 41 465 4.0 1616 2 Q9kka3
 42 465 4.0 1617 2 Q9kka3
 43 464 4.0 1265 2 Q9fda0
 44 463.5 4.0 1616 2 Q9kba0
 45 463.5 4.0 2349 2 P94750

Thu Sep 13 14:18:21 2001

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "the genome sequence of the plant pathogen Xylella fastidiosa.";

DR Nature 406:151-159(2000).

RL EMBL: AE003982; AAF84338.1; -

SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5DD24E10 CRC64;

Query Match	10.5%	Score 1222.5;	DB 2;	Length 2059;
Best Local Similarity	24.0%	Pred. No. 3.6e-43;		
Matches 599;	Conservative 368;	Mismatches 855;	Indels 679;	Gaps 115;
QY	39	GSSPVIRLTRVATLAILVIGATLNGSAYAQNNKSIARCTTGNNDNASASH-EASIAIGSL	97	
DB	13	GKEPLHTILFVALSLSLTPYTEKANAQYI--NDGMDGCGORIVDNGSASGVERTVA---	66	
QY	98	AKAHANQATAIGSK-PDPNCAANOKAGSHAKGKESIAIGDGYLAEGDASIASIGDDLY	156	
DB	67	TOCESDPTVTGYARFFGSPGTAAEQOAS-----RNLTGGSLYVN---SGQGVNDV-	117	
QY	157	LDRNSTNSKYPNGLLSTLTONHTVLRIQRDSNGSKYRTAAEGHASTAVGA--MAYAKG	214	
DB	118	-----LNKYSIRMGSVITMN-TVA-----GTNAIIGSQAASDAALKASLATRASG	164	
QY	215	HFAAFCTRTAEGNSLAVGLTAKAEKG-YTIAIGSNAQAIFYGALGADTRVDL-DY	272	
DB	165	ARAIAGAKASADGVDTVALGSCATAGTCASSIAIGLNASAVN-GAVAVGGGALVTVPDG	223	
QY	273	GIALGYSQILNNNNNNNNKAYYPEGNSIKSKATGNGL--FSIG---SSTIK-RKII	326	
DB	224	AVALLGLNSVA---STGKGLSGYDPKTKTSTDSAAKWSLTAASVIGDVSNTLKTROLS	280	
QY	327	NVGAGYEDTAVNVAQLKAVENLAKR--QITFFKDDNGT-----GVKKLGEITL	374	
DB	281	GLAAGTSNTDAVNAQLKAVDEITASRNWLTASGANSVAPGSSVDLKNLTITTKA	340	
QY	375	IKGGGTQ-----ADKLTDNNIGVVD---NNTCLKVKLAKNLSGLETVST	417	
DB	341	IGSDVQFNLNKLKDKVITTLAVGDALLNTDGIAGTDVLSLTGLAI-----TDGPAVTA	394	
QY	418	KNLTASEKTVYVSGNNTAELOSGGLTFTPTTNASTDKTYVGTDLKFTDMSNTALEDTTR	477	
DB	395	SGIDAGSKV-----ISHVAAGAVSET-----STDAVNGSQLNAVVOVASQP	435	
QY	478	ITKDKITGFSNKAGTVDENKPYLDKDKLVGNSTLNGSLTVNNTIGGSN-KQIQVGADGI	536	
DB	436	VT-----FTGNEGAV-----KRLGQSVVISGESSTAGTYSGNLKSVVDEAAG-	479	
QY	537	KFADVNVNNAKFCGTTTRITEEIEGCFADAGDKVDKSPYLDKQLQGVGVKTIKDSGIN	596	
DB	480	-----RIHLQADSPKFCGVNIN-----GG-----KISGVT	506	
QY	597	AGDKQISNVKDATDDTDVATYKQLKQVQ-----QADG	629	
DB	507	AG-----TEETDAVNFSQLKSISTAVDQGTTLTASGNSGKVASGGTVDLKNPTDG	556	
QY	630	ALQSFISIRDEKGOFTISNLYS-NGNTPNFTETITFAGENGISINDIAKGVKVGIDPI	688	
DB	557	NLTISKSGSDNVVFNLSKDFKVDGMTSGT-----TVVNDGVKGVSDVALGTITGIDG	612	
QY	689	NGLTTPKLTGSDKDGKTLQVIEQVAGS---NDTKIIRGLSPTLPSITNAGGVRTTQGG	745	
DB	613	PAVTASGIDAGSK-----VISHVAAGVYVSETSTDAVNG-----SQLNAVVOVAS-QP	658	
QY	746	NTITSDEKSKAASIGDILNTGFLNKNNSVGFVSTYNTVDFIDGNATTAKTYDETQO	805	
DB	659	VFTGNEGAVKRS-----LGOSVVISGESSTAGMYSGNLKSVVDEAAGRIHLQADSPK	713	
QY	806	TSKVYDVNVDEKTIETLTCGDKGNTKI-----GVKTT-----TLTITNANGKATNFSTT	854	
DB	714	FGNVV---INNGKISGVTAGTEETDAVNFSQLKSISTAVDQGTTLTASGNSG-----SKV	766	
QY	855	DNDALVNARDIAENLTAKELHTTKGTADTALQTFKVKKGATVDETIIVTGKDGTONGK	914	

DB	767	ASGGTVDLKNTDGNLT-----ISKSGSDSNVVFVFNLSKDFKVDGM	805						
QY	915	TVNTLKLKGENGLVATNKNKDGVTGINTQSLKAGDSTTLNKDGLSTIKNPASNEQIQVG	974						
DB	806	T-----SGTTVVNN-----DGVKGVSDVALGTGLTI-----	832						
QY	975	ADGVKFAKVDKGNSSGIDGTSRIITKDQIGFTGANGSLDITTKPHLTKDKLKVGEVEITNT	1034						
DB	833	-----TDGPA-----VTAS	841						
QY	1035	GINAGGKRIITNIOGDIOTNSNDVGTGRVYDLKTELESKINSAAKTAQNSLHEFSVADE	1094						
DB	842	GIDAGSVISHVAAGVSETSTDAVNG-----SQLNAVQVA-----	878						
QY	1095	OGNIETVSNPYSSYDTSKTSVDITFAGENGITTKVKNKVVYVIGIDOTKGLTTPKLTGVNN	1154						
DB	879	-----SQPVFTG-----NEGAVKRLSGSVVISGESSTAGTY	911						
QY	1155	NGKGI--VIDSKDQO-NFITGLSNTLANVTNDGAGHALSOGLANDTKTRAA-----SI	1205						
DB	912	SGNLKSVVDEAAGRIHLQADSPKFGNVVINGGKI--SGVTAGTEETDAVNFSQLKSI	969						
QY	1206	GDVLNAGENLOGNEAVDFVSTYDVF--IDGNATTAKVYDDTSKTSKVYVDV--	1259						
DB	970	STAVDOGWTLTASGANGSKVASGGTVDLKNTDGNLTISK-----SGSDNDVFNLSKDFK	1024						
QY	1260	-----VDNKTEVTSKKLGVKTTTLTKTSANGNATKFSAADGDALVKASDIATH	1309						
DB	1025	VDGMTSGTTVWVNDGVKGVSDVALG--TTGL--TTANGPAVTASSIDA-----GSKVISH	1075						
QY	1310	LNTLAGDIOTAGCASOASSASVVDADGNKVIYDSTDKKYOVNDKGOVDKNKEVAKDKL	1369						
DB	1076	V-----AGAVSETSTDAVNGSQLNAVQVQASQPVFTTGNB-GAV--KRLGQSVV	1123						
QY	1370	VAAQOTPDGTLAQMNVKSVINKE-----OVNDANKKQGINEDN-AFIKGLENAAKDTKT	1422						
DB	1124	ISGESSTAGTYSGNLKSVVDEAAGRIHLQADSPKFGNVVINGGKISGVTAGTEED-	1182						
QY	1423	KNAVTVGDLNVAQOTPLTFAGDTGTTAKKGLTETLTKG-----GOTDNTKLTDNNI	1474						
DB	1183	---AVNFSQLKSI-----STAVDOGWTLTASGANGSKVASGGTVDLKNTDGNL	1227						
QY	1475	GVVAGTQGTFTV--KLAKDLNLSVYVAGGTTRIDKGISF-----VDANGQAKANTPVLSA	1527						
DB	1228	TISKSGSDNVVFNLSKDF-KVDGMTSGTTVWVNDGVKGVSDVALGTGLTIANGPAVTA	1286						
QY	1528	NCLDLGGKRISNIGA-AVDD--NDAVNFKQFNEVAKTVNNLNQSNQSGASLPEVVTDANG	1584						
DB	1287	SGIDAGSKVISHVAAGAVSETSTDAVNGSQLNAV-----QVQASQPVFTT-----	1331						
QY	1585	KPINTGDKPKQKAI-----KGADKYYHANANGVPVDPKDGK---PITDADKLANLA	1632						
DB	1332	----GNEGAVKRLSGSVVISGESSTAGTYSGNLKSVDENAGTIHLQADSPKFGNVV	1387						
QY	1633	AHKPLDAGHQVVASLGG--NSDAITLTLNLIKSTLPIQDTPNTGNANAGQASLPSLSAAQ	1690						
DB	1388	-----INNGKISGVTAGTEETDAVNFSQLKSISTAVD-----QGWTL-TASGAN	1431						
QY	1691	QSNAA--VKDVLNCGFNLOTHNHQVDFVKYDVTVFNVTGTCADITSVRSADGTMNITV	1748						
DB	1432	GSKVASGGTVDLKNTDGNLTISK-----SGSDNDVFNLSKDFKVEKSTIV	1476						
QY	1749	-NTALAATDDGHNVLKAKDGKEYKADDLMPNGSLKAGKASDAKT-PTGLSLVNPNAKG	1806						
DB	1477	GNTOL---DKDG---VKVSSNVLLDSNELV-----ITSHISTSVKTLANGESVNVTVV	1526						
QY	1807	G---STGDVALNLSKAVFKSKDGTITTTTVSSDGIISQCKDNSSITISKQGLNVGGKVI	1863						
DB	1527	GDGVNIDDDVVVNDL-----GLSIVG--GASLTL--GINAGSHKI	1563						
QY	1864	SNVGKGTGKTDAAVNOQLNEVRNLLGLG-NAGDNDAGNOV---NIADIKK-DPN---SG	1915						
DB	1564	TNVTAGTETDVAVNFSQLKSVSEAVDKGWTLTASGANGSKVASGGTVDLKNTDGNLAI	1623						

Thu Sep 13 14:18:21 2001

QY 986 -----GNSSTGIDGTSR----- 997
Db 1113 DGESBETQEVKAGDKVTFKAGKNLVKASEKDFYSLQDTLTGLTSITLAGTANGRND 1172
QY 998 -----ITKQIGFTGANGSLDTPKPHLKVKGEVEITWTGKAGGKKTNIQSGDITQ 1053
Db 1173 TGTIVKDKGLTITLANGAACT-----DASNGNTISVTKDGISGAKETITVSKA--- 1222
QY 1054 NSNDVITGRVYDLTELESKINSAKTAQNSLHEFSVADEQGHFTVSNPYSSYDTSKT 1113
Db 1223 -----LKYKDTQ-NIADETDQKEFHA-----AVKN 1247
QY 1114 SDVITFAGENGITTKV---NKGVRVIGIDQTK-----GLTTP-----KLVGNNGKGV 1160
Db 1248 ANEVFVGNGATVSAKTDNNGKHTVTDVAEAKVGDGLEKDTGKTKLVKVDNTDGNLL 1307
QY 1161 -IDSKDQNTITGLSN-----TLANVTN-DGAGHALSQG-----LANDTDKTRAASIGDV 1208
Db 1308 TVDATKASVAKGEFNAVTTDAITAGTNANERGVVGVKSGNGATATETDKKKVATVGDV 1367
QY 1209 LNAENLQNGEAVDFVSTYDVTDFIDGNATTAKVYDDT-----SKTSKVYDVVNDK 1263
Db 1368 AKK-----INDATEFKVENDDSATIDSPDSDGANDALKAGD 1405
QY 1264 TIEVTSKKLVKTTTLTKTSANGNATKFSADGDALVKASDIATHLNLTAGDIOTAKGA 1323
Db 1406 TLTAKGNLKVK-----RQGNITFALAN-DLSVKATVSKLS----- 1444
QY 1324 SOASSASYVDADGNKVIYDSTDKKYYVNDKQVDKNEKVAKDLVACAOPTDGTLAQM 1383
Db 1445 -----LGTNGKNVNTSDTKGLNFAKD-----SKTGDANIHLNGIASLTDTLLNS 1491
QY 1384 NVKSVINKEQVNDANKOGINEDNAF-----IKGLENAADTKTKNAAVTVGDLNAVOT 1438
Db 1492 GATTNLGGNGITDNEKRAASVKDVLNAGNVGVKSPASANNQVENI-----DFVATYDT 1546
QY 1439 PLTFAGDTGTTAKKLGELTTLTKGGQDTNLTNNIG-----VWAGTDG---FTVKLAKDLT 1492
Db 1547 VDFVSGDKDTT-----SVTVE---SKDNGKRTKVIKAKTSVIKDHNGKLTCKELKDN 1598
QY 1493 NLNSVNAVGTIDEKISGFVANDGQAKANTPVLISANGLDL---GGKRISNIGRAVDDND 1549
Db 1599 N-----NGVTTETDGRDEGNLVTAKAVIDAVNKAAGVKTGTGANGQND-- 1644
QY 1550 VNEKQFNEVAKTVNNLNOSGASLPFVVTDAKNGKPTNGDKPQKAIGA-DGKYIYA 1608
Db 1645 -----FATVASGTN-----VTFADG---NGTAEVTKANDGSITVKYVK 1681
QY 1609 NANGVPVDKGPITTDADKLANLAHKG---PLDAGHQVVASLGSNSDAITLTIKSTLP 1665
Db 1682 VADGLKLDGD-KIVADTTVLT--VADGKVTAPNNGDGKFEVDASGLADALNKLSTAT-- 1736
QY 1666 QIDTPNTGNAGAOGLPSLSAAQSNASVKDVLNNGENLQTNHNOVD-----VK 1718
Db 1737 -----AGKEGTGEVD--PANSAGQEVKAGD-KVTFKAGDNLKIKQSGDFTYSLKKEK 1787
QY 1719 AYDTVNFVN---GTGADITSVRSADGTMNSITV---NTALAATDDGNVLIAKAKGKYK 1772
Db 1788 DLTSVEFKDANGTGSESTKI-----TKDGLTITPANGAGAAGANTANTISTVKD-- 1838
QY 1773 ADDLMPNGLKAGKASADKPTGLSLVNPNGAKSGTGDAVALNNLSKAFVKS-----K 1826
Db 1839 -----ISAGNKA-----VTNVVSGLKFKFGDGHTLANGTADFEKHYDNAYK 1879
QY 1827 DGTFTTTVSSDGIISQKDNSSITLSKDLNVG-----GKVIS---NVGKGTDDTAANY 1878
Db 1880 DLT-----NLDEKADGNPTVADNTAATVGLRGLGVWISADKTTGEPNOEYN-AQV 1930
QY 1879 QOLNEVRNLLGLG-NAGNDNADGNQVNIADIKDP-----NSGSSSNRTVIRKAGTV 1928
Db 1931 RNANEVKKSGNGINVSGKTLNGTRITFELAKGEVVKSNFETVKYNADGSETNLVKVGD 1990

QY 1929 LGKGKND--TEKATGGVGVGDKNANGDLSNVWVKTKDGSKKALLATYNAAGOTN 1986
Db 1991 YYSKEDIDPATSPKMTG-----KTEKVKVNGKV-----VSANGSKTEVTLTNKSG-- 2037
QY 1987 YLTNNPAEADIRINEOGIRFFHVNDQNGPEVVGNGRIDSSASGKHSVAIGFOAK----- 2041
Db 2038 YVTGN--QVADAIKAKSFEL-----GLADAAEAKAFASAKDKQLSKD 2079
QY 2042 -----ADGEAAVAIGROTQAGNOGTAIGDQASIAITGTVNVTGKHSAGIDPS 2095
Db 2080 KAETVNAHDKVRFANGLNTKV---SAATVESTDANGDK-----VTT-----T 2118
QY 2096 TVKADNSYSVGNNOQFIDATQTDVFGV---GNNTITVT-----ESNSVALGNSAISAG- 2145
Db 2119 FVKTD-----VELPLTQIYNTDANGNKIVKADGKWTYELNADGTASNKVETLGN 2167
QY 2146 --THAGTQAKKSDGTAGTTTTTAGATGTVKFGAGOTAVGAVSVG-----ASGAE 2191
Db 2168 VDANGKKVYKVTENGADKWKYTNADGAADTKGEVSNKYSTDEKHVRLDPNNQNGK 2227
QY 2192 RRTONVAGEVSATSTDAVNGSOLYKATQGTAN--ATNELDRIHONENKANAGISSAM 2248
Db 2228 VVIDNVANGEISATSTDAINGSQLYAVAKVTNLGAGQVNNLEGVKNVGRADAGTASAL 2287
QY 2249 AMASMPQAVIPGRSVMTGGIATHNGSGAVAVGLSKLSKDSQGWFKINGSADTQGHVGA 2308
Db 2288 AASQLPOATMPKSMVAIAGSYQSQGLAIGVSRISDNGKVIIRLSGTTNSOGKTVGAA 2347
QY 2309 GAGFHF 2314
Db 2348 GVGQW 2353
RESULT 3
ID Q9F3X5 PRELIMINARY; PRT: 2712 AA.
AC Q9F3X5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE MAPB PROTEIN.
GN MAPB.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277636; CAC14203.1; --
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;
Query Match 8.9%; Score 1037; DB 2; Length 2712;
Best Local Similarity 23.1%; Pred. No. 3.1e-35;
Matches 636; Conservative 312; Mismatches 935; Indels 868; Gaps 132;
QY 12 ATGTFMAVEACAKSHSGSSSTAGQVSSPVRLTRVATL--AILVIGATINGSAYQN 69
Db 374 ASDNSIALGNAARTNKDSDSIAIGSAETKAAHSLAVGTTSKALAEALALGKLAEGK 433
QY 70 NSKTAFGTT-----GNNDNASASNEASIAIGSLAKAHANQATIG-----GSKP 113
Db 434 TSSVAMGNTSKADGNSVAVGNTSQTLOS--TIAIGSSAIANPERTISIGLNACKGQEA 491
QY 114 DPRNOAANQ-----KASHAKGKRESTAIGDVLAE--GDASTAIGSDDLYLDNRNSTNSKY 166
Db 492 DATGTKHQSQINIGNSGEGVIGQLNIGIGHAGTNVVGKHNIALGT---YAGTNLKNSEE 548
QY 167 PNLGLSTLIQHTVLRQIRDSNGSQYRRTAAEGHASTAVGMAYAKGHFANNAFGRSTA 226

Thu Sep 13 14:18:21 2001

QY	1968	KDGSKALLATYNAAGTNTLTNNPBAIDRINEQIRFEHVDGNBPVVOGRNGIDSS	2027
Db	2443	KDGDVAISGKDGVG-TIGLT-CPAGA-DCKNANAI--IGVNDs-----VKGLDGNCK	2492
QY	2028	ASGKHSUVAIGQAKADGAEAAVAGROTCAGNQSTAIGDQAQATGDOSIAIGTGNVTKG-	2086
Db	2493	-DGNsKTRIV--TKPNGBEE-----EQVATNDGLVFG--ADKGEHKAKRLGTVKVGDD	2543
QY	2087	---HSGAIGDPSTVKAADRSYSGVNNNOFIDATQTDVFGVGNNTVTVESVALGSNSAIS	2143
Db	2544	KNTEVEAGDTIRVLKONIDVKGIN-----VTENLVKEGAKINGNV---2588	
QY	2144	AGTHAGTQAKSDGTAGTTTGTAGTGTGKGFAGTAVGAVSVGASGAERIQNVAAGEVS	2203
Db	2589	-----IDGVADEVN	2598
QY	2204	ATSTDAVNGSOLYKATQGI---ANATNELDRIHQENKANAGISSAMAMSPQAYIPG	2260
Db	2599	ATSKQAVNGSOLHVKQOVNQNAITANKLDGHINKVDKDLRAGIAGATAVAFLOPNEAG	2658
QY	2261	RSMTGGIATHNGGAVAGVAKSLDSNGQWFKINGSADTQGHV--GAAGV	2309
Db	2659	KSIVSLGVGSYRSALATGAVARNSDNNKISIKLGGMNSRGDVNFGSGIG	2709
RESULT	4		
Q9PC04		PRELIMINARY:	PRT; 1190 AA.
ID	Q9PC04		
AC	Q9PC04		
DT	01-OCT-2000	(Tremblrel. 15, Created)	
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)	
DE	01-MAR-2001	(Tremblrel. 16, Last annotation update)	
DE	SURFACE PROTEIN.		
GN	Xf1981.		
OS	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		
OC	Xylella		
OX	NCBI_TaxID=2371;		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=9A5C;		
RC	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvaranga R., Alves L.M.C., Arya J.E., Baia G.S., Baptista C.S.,		
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,		
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,		
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,		
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Mench C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,		
RA	de Rosa A.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,		
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,		
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,		
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,		
RA	Vallada H., van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,		
RA	Zago M.A., Zatz M., Weidman J., Setubal J.C.;		
RL	Nature 406:151-159(2000)		
DR	EMBL: AE004017; AAF84783.1;		
SEQUENCE	1190 AA; 118446 MW; 756741B0C8D787CC		CRC64;

Query Match	6.4%;	Score 743.5;	DB 2;	Length 1190;
Best Local Similarity	24.1%;	Pred. No. 2e-23;		
Matches 352;	Conservative 188;	Mismatches 481;	Indels 441;	Gaps 67;
QY	1010	GSUDTTKPHLTDKLKVGE---VEITNTGINAGGKKTITNQSDITQNSDNDVATGGR--	1063	
Db	13	GGIEGVDP-LFLPKYKIGQSLQLOHAVMTSA--ASSKK-----CTQPRRNNNAWAKRSA	62	
QY	1064	-VYD-----LKTELESKINSAKTA-----QNSLHEFSVADQEGNHFTVSNPY	1105	
Db	63	KLHRRRLHVLVLLTAAASTGYTKVAAQVYVNSDSTENCVEILGDSSTQTSFIHASNDK	122	
QY	1106	SSYDTSKTSVDITFAGEN---GITTKVNGVVRVIGDQTKGLT-TPKL-TVGNNGK---	1157	
Db	123	CKPDETQTEVSLFYDYRNVLGGLVYNBEG--KLGLVDISGATYSMLRGLSIATMNGSAGI	180	
QY	1158	-GIVIDSKDQONTITGLSNTLANVTNDG--AGHALSQGLANDTKTRAISIGDVLNAGFN	1214	
Db	181	DSIAIGSGGSGKT-----DQNTSGATVAQGL-----RSIAIGTTARS---	217	
QY	1215	LOGNCEAVDFSVTYDVFIDGNATTAKVYDDTSTKTSKVYVDVNDVNDKTIETVSDKILG	1274	
Db	218	-----QSQDAISIGTGASTGTGNFAI-----237		
QY	1275	VKTTLTKTSANGNATKESAADGALVKASDIATHLNLTLAGDITQAKGASQASSASYVD	1334	
Db	238	-----AINGALATSIANGIALGASSSVTTR-----GGVALGQGSAAATASGI---	279	
QY	1335	ADGNKVIYDSTDYKYQVNDKGVQDNKEVAKDKLVAQAOTPDGTLAQMNVKSYINKEQV	1394	
Db	280	-----TGYPDVTKS-----TSTLSVSMWRSTLGAVSI	306	
QY	1395	NDANKKOGINEDNAFTKGLNENAAKTKTKNAATVGDNLNAVAQT-----PLTFAGDTGTT	1449	
Db	307	GNITSS---TSQTRQLTGLAAGRSDDID---AVNVAQLKLLAESYGGGWNLTASGSSN	359	
QY	1450	AKLIGETLITKGGOTDITNKLTNNIGVAGTGG--FTVKLAKDLNLNLSNAGGTIDEX	1507	
Db	360	V-ALGESVDLKN-----SDGNLLITKTTDNDVTFNLATAL-KVDSLTTGNTAMTDD	409	
QY	1508	GISF-----VDANGOAKANTPVLSANGLDLGGKRISNIGANVDNDVNFKOFNEVAKTV	1562	
Db	410	GVTVGKRVTLSTGLVIAEGPSVLISSGAINAGQIMNVGTGTADTDAVNFQGLQAVSDTA	469	
QY	1563	NNLNQNSGASLPVVTDDANGKPICTDGGKPKQKAIKGADKYHYHANANGVPVDKDKPI	1622	
Db	470	SKGWNLLASGTN-----SSNVAPGASVD-----LAKTDGNLLITKAIGI-----	508	
QY	1623	TDADKLANLAHAGKPLDAGHGVAVSLGNSDAITLITNISTLPQIDTPNTGNA-NAGOAQ	1681	
Db	509	--NDVTFNLAT---ALEADSLTTGNTAMTDDGVTVGS-----NVTLSGLVITDGPVS	557	
QY	1682	SLPSLSAQAQ---SNRAISKVYDLNNGFNLOTNHNQVDFVKAYDTVNF-----VNGTGADI	1733	
Db	558	TSGISAGNOKIITNVAAGTADT-----DAVNFSQLQAVSSSTASKG	597	
QY	1734	TSVRSADGTVNSITVTNTALAAATDDGCVLI-----KAKDGKEYKADDLMPNGLSKAGKAS	1789	
Db	598	WNLASGANSNVVPGESVDLKNSDGNLLITKTTDSDNDVTFNLATALKVD-SLITGNT--	654	
QY	1790	DAKPTPTGSLVNPAGSGTSDAVALNNLSKAVFKSKDGTGTTTTVSSDGIISQGDKNSSI	1849	
Db	655	-AMITDGVTVGSNVTLGSTG-----LVITDGPVS-----682		
QY	1850	TLSKDLNAGVKVTSNVGKGTGKTDAAVQOLNEVR-----NLLGLG-NAGN-----	1895	
Db	683	--TSSGISAGNOKITNVAAGTADTDAVNFSQLQAVSSSTASKGNLLASGANSNVAPGES	740	
QY	1896	---DNADGNVNIADIKKDPNDSG-----SSNRTVTKAGTVLGG	1931	
Db	741	VDLKNITDGNIV-----ISKESGNDVLFNLSSSLKDLKLTGVTVMVTGTVGSGVTLGS	796	

[illegible]

```

RESULT      5
Q9F285      PRELIMINARY;          PRT;  3705 AA.
ID          Q9F285
AC          Q9F285;
DT          01-MAR-2001 (TrEMBLrel. 16, Created)
DT          01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT          01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE          YAPH PROTEIN.
GN          YAPH.
OS          Yersinia pestis.
OC          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC          Yersinia.
OX          NCBI_TaxID=632;
FN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=CO-92 BIOVAR ORIENTALIS;
RA          Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT          "Evolutionary origins of the autotransporter proteins.";
RL          Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR          EMBL; AJ277631; CAC14227.1; -.
SQ          SEQUENCE 3705 AA; 370756 MW;  714FDF16455968C9 CRC64;

```

Qy	113	PDRPQNAQKAGSHAKGKESTIAIGDVLAEQDASIAIGSDDLYLDRNSTNSKYPNGLLS	172
Db	181	IDFLGSSQLIYMGAYGAATNSITWTFGDIL-----NDVVNDRAQIEGIVNKLAF	229
Qy	173	T--LIQNHVTLRQIRDSNGSQRYRTAAABGHASTAVGAMAYAKGFANAFCTRSTAESNY	230
Db	230	TGRPHVHT-----GSSVTSFVSTGGANNFTST-----MDFASGADVKIDRTGSTGD---	275
Qy	231	SLAVGLTAKAEKGYTTAIGNSAQAINYGALALGADTRVDLDYGI-----ALGYSQILNN	285
Db	276	LTSTGYNAFA--YTFADGASPELIANQVSGFTTNRGLEIGSYNSIDGFGSGKVLVQ	332
Qy	286	NNNNNKAYVPENGSGNIIKSSKATG-----NGLPSIGSSSTIKRKIINVAGYEDTD	336
Db	333	SRSDGS---IISGNGIDNATTNAAAGINNNASGDANVIYNLGTGSI-LKATNTG-----	381
Qy	337	AVNVAQLKAVENLAKROIITFKGD-----DNGTGVKKKLGETLTIKGGETQADKLT	386
Db	382	---ILATKNANNASDIYIRAGDITAAATGISTATHNGTG-----TVKIK-----	421
Qy	387	DNNNIGVVTDNNTGLKVKLAKNLGSLGETYSTKNLTASEKVTGVSGNNTAEQLSGGLT-FT	445
Db	422	--NDGTFITSTAGIAISSA---SIKEISVDNTDGTITATAGTVNV--LASAILNLFG	472
Qy	446	PTTNASPKDVTYGTDLKFE--TDNSNTALEDDTTRITKOKIGFSKNKAGTVDENKPYLDKX	503
Db	473	GTINTSAT-----ANGITTAGTEGGHTLDLRINLLGTGIALSNVAGV-----N	516
Qy	504	LKVCNSTLN--NG-----GLTVNNTIGGSNKQIQVGADGIKPADVNVVNSNAKEGTT	554
Db	517	LTSNVPLNTLNGTALNSLTGTLFVDSLGRN-TNIEGAGIGIAAANT-----	564
Qy	555	RITEEEIGFADAGKVDKPSYLDKKQLQGVGWKI---TKDSGIN-AGDQKISNV-----	605
Db	565	-----ELNNTDAAE-----LD---INVNAGIGIQATGGVNLASNLIINVANTLG	608
Qy	606	-----KDATDDTDAVYKOLKQVQODADGALQSFSIRDEKQOEFTISMLYNGNTPNTE	660
Db	609	TALQITDGIDNTTIG---NEIQLNAENAT-----AINFLGSSKTLNNG	651
Qy	661	TI---TFAGENGISINDIAKGVKVGIDPINGLETTPKLTVGSDKDKGTOLVIEQVASG	716
Db	652	TIKGSVIFAGVADHIINN-----NGTLDGTLTGA---GNDTLFVLDDSSQS	694
Qy	717	NDTKNIIRG-LSPTPLSPITNAGVRRTEQGNITISDEKSKAASIGDT-LNTGFNLKNS	774
Db	695	NDVNLGDGNSVTIQNGATVSSITGNGDNDFTINGMSVSTYLGSLDAGTGLNTYXX	754
Qy	775	NSV-----GFVSTYNTVD-----PIDGNATTAKVTYDETNO-----	805
Db	755	ASTDELAATSLQGF-TNINLVDSHITLVSDNIGSGVMNIDSSSELLFGSTFGDILHAT	813
Qy	806	-----TSKYTYDVNDEKTIELTGNGKTNKIGVKYTTTLTTNNAN--GKA-----	853
Db	814	LGAGTGSIAIVNNSANYSLEQASMFAGTQVWNOGG---ALTAASNQLGSAKIGLDGTLN	869
Qy	854	TDNDALVN-----AKDAENLNTLAKIEHTTKGTADTAL-----QTFVKYKDA---	897
Db	870	LDNTALEHNVLGTNGTLNVAKNLATTFPFGVTVGAGSGVIVNLTKTTFALSADNAALA	929
Qy	898	-----TDDETTIVGDKGTQNGKVTNVLKLGNGLTVATNKDKGVTFE-GINTQ---	947
Db	930	SATLKLSDDSVTVVG---TTDRTLHGLDSG-----GTLIFDGAVPQSQTSGV	974
Qy	948	KAGDSTTLNKDGLSIRKNPASNEQIQVAGDGVKFAKVDKGNSTGIDGTSRITKQIGFTG	1007
Db	975	VTVTDALNSGTVNITGSGSWONTDPLATNVSLIEQDRAGS-----TLELINATN	1024
Qy	1008	ANGSLDITTKPHLTKDKLKYCEVEITWTGINAGKKITINIQSGDITONSNDVATGGRVYL	1067
Db	1025	VTGDIDAL-----DLLVNGTATISG-----TQG-----	1047
Qy	1068	KTELESKINSAAKTAQNSLHEFSVADEQGNHFTVSNPXYSDTSKT-----SDVITFAG	1121

Db	1048	--VQSAIQGGSTVANAIHNYGLASSNSN--GDSGLVYVNTLSALELLADGADALLAT	1102
Qy	1122	ENGITTK--VNRGVVRVG---IDQTKG-----LTPKPLTVGNNN--KRG	1158
Db	1103	ESGLTANRLNAELFGVGLVYDAONGALTLANGSNRYEGTTVTAGELILGANGAFQOT	1162
Qy	1159	IVTDSQGN--ITGLSNTLANVND-----GAGHALSQGLDMDTKRAASTGDVLNAG	1212
Db	1163	SLLDIAGSASANTNGYSQVAGVNTVCIVTLGSGGLTSGLLTN-----GGILDUTGGA	1216
Qy	1213	FNLOGGGEA--VDFVSTYDVPFDGN-----ATTAKVYDDT--SK	1250
Db	1217	LNLTGXCSTVAGLTGAGTLNGLNGLSVSAANGLSGQTHIADVASVTLTDGTGLTS	1276
Qy	1251	TSKVYVYVNDKTIETVSDKLGCVKTTTLTKTSANGNAKFSAA-----DGD-	1298
Db	1277	AVEVLGTNLUNGANAMTVNLSCDGTINTNAAVTLGSN-NSFSGAHQIGTDELTVGQAS	1335
Qy	1299	-----ALVKASDIATHL-----NTLAG-----DI-----OTAKASQASSSASY	1332
Db	1336	NLGASSTVNLGTLTSHLLNGVSESIANVLSVAGSVTDIIGGADTALTANNSGFLQY	1395
Qy	1333	VDADGNKYIDSTD-----KYYQVNDKGVDK	1360
Db	1396	ALAGNSKLTVASTNNILGASSSVAGAGDTSLSGFNGTGFSGVTLQVTDDAEV--	1453
Qy	1361	NKEVAKDKLVAQOQTPDGTLOAMNVKSVINKEQV-----NDA-----	1397
Db	1454	-TLTSSNGVSNVAVIDADATLNDLDALENHVLTCNGLLNVAKNDASTAFDEGTVGG	1511
Qy	1398	-----NKKOGINEONA-----FIKLE-----NAA	1417
Db	1512	AFSGIVNLNTTFALSAONAAARATLKLSDSVTVTGATDRTLHGLDLNGTLTFDGS	1571
Qy	1418	KDTRKNAAVTVGDLNVAQT--PLTFAG-----DTGTAKKILGETLT	1458
Db	1572	PPOSQANGVTVTDLALNSGTISITGAGNWEHPVTPPNVSLLEQDRGDLLELINAAN	1631
Qy	1459	IKGGOTDINKLTDNNTCVVAGTDG-----F	1483
Db	1632	VTGANNLDLLVGT-AITSGTQGVESAIOQGGSTVANAIHNYGLTSSNGGSGELVNY	1690
Qy	1484	TVKIAKDLTN-----LNS--VNAGSTRIDKGISFVDANGAK-----	1519
Db	1691	TLSALELLANGANALLATESGLTANRLNAELFGVGLVYDAONGALTLANGNRVEGT	1750
Qy	1520	-----ANTPVLsangldlGGRKRSNI--GAAYDDNDVANKFQ--FNEVARTVNN--	1564
Db	1751	TTVTAGELILGANGA-FQOTSLNIASGASANTN---GYRQIVGAVTNSGAVTLNGGVL	1806
Qy	1565	-----LNNQSNSGASLPPVVTVDANKPRTDGTGPKOKAIKGAQKYYH	1607
Db	1807	TSGLLTNGIILDTGGLNALAAGSSTVAGLTGAGTLNNGDGLAVSATNSGLSGQTHI	1866
Qy	1608	ANANGVPVDKQKPIITDA-DKLANLAHKPLDAGHGVASLGG--NSDAITLT--NIK	1661
Db	1867	ADVASTLTGTGLTGSFAVEVLGTNLUNG--ANAAMTNVLSSGGVINTNAAVTLSSNNSF	1924
Qy	1662	STLPQIDTPTNGANAGAOQSLPSLSAAQSN-----NASVKD	1699
Db	1925	SGAHQIGT--DGELTVGQASNLGASSATVNLGTLTSHLILNGVSESIANVLSVAGSVTD	1982
Qy	1700	VLANVGNLQTNHQDFVKAYD-----TVNFVNGTGA-----DITSVRADGT	1742
Db	1983	IIG-GADTALTANNSGFLQYALAGNSKLTVASTNNILGASSSVAGAGDTSLSGFNXT	2041
Qy	1743	MSN--ITVNTALATDDD-----GN-VLIKAKQKFTKAD-----DLMPSGLKA	1784
Db	2042	FCNSVTGSGVLOVTDAAEVLTTSSNVGNTVKVDIADATLYVNDIALDLHDVLTENGTLN	2101
Qy	1785	GK-----SASDAKTPTGL---SLYNPNAGKSGTGDVAALNN---LSKAFKSKDGTFTT--	1832

Db	2102	AKYLATPDXGSTVGXXFSGIVNLT	-----NTTEALSDADNAALARATLKLSDSDSVTTVG	2157	
Qy	1833	-----TVSSDGLISIQKDNSSITLTKDGLNVGKVISNVGKTKD	-----1874		
Db	2158	TTDRIHLGLDLNGGLTFIDFGPPQSQANGVYVTDIALNSG	-TISITGACNWEHEPVT	2215	
Qy	1875	AANVQQLNEVRN	-LLGLGNAGNDADGNVNADIKKDPNNGSSSNRTVIKAG	-----1926	
Db	2216	PPNVSLLEQDRGDIILLQLLDADNVTGNANDELMELMINTTISAGQGVQSTVOOGGYTVANA	2275		
Qy	1927	-----TVLGGKG	-----NNDTEKLTG	-----GVQGVCDKDGNAAGDLSNV	-----1962
Db	2276	THNYGMTSNGSGLYVNYVLTSLALELLADGANALLATATESG	TANRELNALNELSGVGLVVD	2335	
Qy	1963	-----VWTKQDGSKK	-----ALLATYNAAGQTNLT	-----NNPAEADRI	1999
Db	2336	AQNGALTLANGNRYEGT	TTVAGELILGANGAFQTSLLNIASGASANINGYROTVGAV	2395	
Qy	2000	NEQG	-----IRFHVNDGNQEPVVGGRNGIDSSASGKHSVAIGFQAK	2041	
Db	2396	TNTGTVTLNGGELTSDTLINTGMINVTG	-----ILMLENGGASSISG	-GLTGTGILNI	2450
Qy	2042	ADGEAAVAIGRQTAQNGQSTAGIDNAQAT	-GQDSIAIGTGNVTVTKHSGAIGDPSTVKAD	2100	
Db	2451	KGGDETIISDNLNLAGQTNLS	-DGRSVTLNGGTLIGTGNLGSV	-IDVLGDLNLV	-AD 2506
Qy	2101	NSYSGVNNNQFIDATQDVFQGVGNNTVTVESVALGNSAISACTHAGTQAKKSDCTAG	2160		
Db	2507	NSLA	-----NVISGDG	-----TINTATVTLSGNSFSFGAHQICTNGELTVCQAS	2551
Qy	2161	TTTTAGATCTVKGFAGQTAAGVAVSVA	-----SGAERRIONVAAGEVSATSTDAVNGS	2213	
Db	2552	NLGASSAT	-----VNLGTLTSHLLINGVSESIANVLISG	-VAGSTVDIIIGA	2596
Qy	2214	QLYKATQGLANATNELDRIHONENKA	-----NAGISSAMAMASMPQAYIPGRSMVTGG	2267	
Db	2597	D	-----TALTNANSGELGQYALAGNSKLTVA	STNNLGASSSVALLG	-----TGDITLSLG 2646
Qy	2268	IATHNGQGAVALGSLKSLDNGQ	2289		
Db	2647	FNGTFGNSVTGSGVLQVTDAAE	2668		
RESULT 6					
Q9F2D8	Q9F2D8	PRELIMINARY;	PRT: 1107 AA.		
AC	Q9F2D8:				
DT	01-MAR-2001 (TrEMBLrel. 16, Created)				
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)				
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)				
DE	SAPB PROTEIN.				
GN	SAPB.				
OS	Salmonella typhi.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_TaxID=601;				
RN	[1]				
RC	SOURCE FROM N.A.				
RC	SPRAIN=CT18;				
RA	Henderson I.R., Nataro J.P., Cappello R., Stein C.				
RT	"Evolutionary origins of the autotransporter proteins."				
RL	Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ277623; CAC14217.1;				
SO	SEQUENCE 1107 AA; 113223 MW; F5C7CA651FED51AB CRC64;				

```

Query Match          5.7%  Score 661.5;  DB 2;  Length 1107;  Gaps 58
Best Local Similarity 22.7%;  Pred. No. 5e-20;
Matches 323;  Conservative 194;  Mismatches 476;  Indels 429;
QY 971 IQVGADGVKFAKVDKGN-----SSTGIDGTSR-----ITKDIQTGTCAGCNSLDITTK 1016
db 37 IMYSADALANAGNDGTGSIQNQNTDTATNTTSINNLSNVLTITDALLMDAAAGSTFSASR 96

```



```
QY 1017 PHLTOKLKVGEVEITNTGINAGGKKITNIQSODITONSNDVAVTGRVYDLKTELESKIN 1076
Db 97 -----NGSASITNLAAGTLAADSTDAVNGSOLFTEKVD--- 132
QY 1077 SAAKTAQNSLIEFSVADEQGNHFTVNPYPSSYDTSKTSVITFAGENGITTKVNGVVRV 1136
Db 133 -----ON-----TADITT-----NTNSINQNTDI 152
QY 1137 GIDOTKGLTTPKLTGVNNGKGIVIDSKQNTITGLSNTLANVNDGAGHALSQGLAND 1196
Db 153 ATWTT-----SINLSNSVTTLTD-----ALLWDAASG 181
QY 1197 T-DKTRAASIGVVLN--AGFNLOGNGEAVDFVSTYDTVDFIDGNATKATKYDDTSKTSK 1253
Db 182 TFSASRNGSASKITNLAAGTLAADSTDAVNGSOLFTEKVDQN-----TADITTNTNS 235
QY 1254 VYDVVNDNKTIETVSDKLGKVTTLTKTSA--NGNATKFSAADGDALVKASDIATHLN 1311
Db 236 I--NQNTDIATWTTTINNLSNSVTTLTDALLWDAADSGTFSASRNGSASKITNLA--G 291
QY 1312 TLAGDIQTAKGASOASSASYVDADGNKVITYDSTDKKYQVNDKQVKNKEVAKDKLVA 1371
Db 292 TLAADSTDAVNGSQ-----LYETNQK--VDQNTSAIAD--- 322
QY 1372 QAOTPGTTLAQMNVKSVINKEQVNDANKKQGINE-DNAFTKGLGLENAAKDTTKTNAAVTVG 1430
Db 323 -----INTSITNLSNDLSNWNETTNSF-----SASHGSSTTNKI--- 356
QY 1431 DLNAVAQTPLTFAGDTGTTAKKLGELT--IKGGQT-DTNKLTDDNIGVVAGTDGTVKL 1487
Db 357 -----TNVAAGELUSESTDAVNGSOLFTEKVDQNTDIAANTTNITQN 401
QY 1488 AKDLTNLNS-----VNAAGGTTRIDEKGISFVDANGQAKANTPVLSSANGLDLGG--KRISNIG 1541
Db 402 STAIENLNTSVSDINTSITGLTDNALLWDEDTGAFSANH-----GGSTSKITNVA 451
QY 1542 AAV---DDNDVAVNFQFNEVAKTVNNLNOSNGSASLPFVVTDAKPKIN-GTDGPKPKA 1597
Db 452 AGALSEDSTDAVNGSOLYETNQVD-----QNTSA-----IADINTSITNMGTD-----A 496
QY 1598 IKGADGKYIHANAGVPVVDKGRPIITDAKLANLAAGKPLDAGHQVAVSLGNSDAITL 1657
Db 497 LSWDDDEGAFSASHGT-----SGTNKITNVA--GE-----IAS--DSTDAING 536
QY 1658 TNIKSTLPQIDTPTNTGNANAGQAQSLPSIAAQO-----SNAASVKDVLNVGFNLQTNHN 1712
Db 537 SOLYETNMLISQYN-----ESISOLAGDTSEYIITENGTVK-----YIRTDN 580
QY 1713 QVDFVKAYDTVNVFNGTGADIT-----SVRSADGTMSNTITVNTALA 1753
Db 581 GLEQDAYATGNGATAVGYDAVASGAGCLALGQNSSSSIEGSIAGLGSSTSNRAITTGIR 640
QY 1754 ATD--DDGNVLIRAKDGKFKADDLPNGSLKAGKSASDAKTPGLSLVNPAGKSTG- 1810
Db 641 ETSATSDGVI-----GNTTDRRELLGALSIGTGESYRQIT-----NVADGSEAQ 686
QY 1811 DAVALLNLSKAVFKSKDGTITTTTSSDGISIQGKDSSITPLSKDGLNVGKVISNVGKGT 1870
Db 687 DAVTVRQLQNAI-----GAVTTT-PTKYHYHANSTEEDSLAVGTDLSAMGAKTIVNADAG- 739
QY 1871 KDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPNGSSSSNTKTVIKAGTVLG 1930
Db 740 -----IGICLNPVLVMAIDAINGIAIG-----SNARANHANSIAM 772
QY 1931 KGNNDPTEKLATGVQGVQVDKDNANGDLSNVVVKTKDKGSKKALLATYNAAGQ----- 1984
Db 773 GNGSQTRGATQDTAYNMDTPQNSVGEFS-----VGSEDGQRO-----ITNVAAGSADPDV 825
QY 1985 -----TNYLTNPAPADIRINEQGRIFRPFHVNDGNOEPVVQGRNGIDSSASGKHSVAIG 2037
Db 826 NVGOLKVTDAQVSRNTQSTINLTQVSNL-----DTRVTNIENGIGDIVTGTSTKY----- 876
```

RESULT 7

```
Q48152 PRELIMINARY; PRT; 1098 AA.
ID Q48152;
AC Q48152;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE ADHESIN (HIA).
GN HIA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NONTYPEABLE STRAIN 11;
RX MEDLINE=96332658; PubMed=8730864;
RA Bareilamp S.J., St Gene J.W. III;
RT "Identification of a second family of high-molecular-weight adhesion
  proteins expressed by non-typable Haemophilus influenzae.";
RL Mol. Microbiol. 19,1215-1223(1996).
DR EMBL; U38617; AAC43721.1; -.
SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;
```

Query Match 5.6%; Score 659.5; DB 2; Length 1098;
Best Local Similarity 22.9%; Pred. No. 5.9e+50;
Matches 323; Conservative 168; Mismatches 457; Indels 461; Gaps 55;

```
QY 1008 ANGSLDPTKPLHTRKDLKV---GEVEITNTGINAGGKKITNIQSODITQN----- 1054
Db 49 ANNTPTV-----NKLKAYGDANFNFTNNSIADAEKQVQVAYKGLNLNKNASDKLLV 102
QY 1055 -SDAVTGGRYDLKTELESKINSAAKTAQNSLHEFSVADEQGNHFTVNPYPSSYDTSKT 1113
Db 103 EDNTAATAVGNLRKLGWLVSSKNGTRNEKSOQVKADEVLFEKGGVQVT---STSENGKH 159
QY 1114 SDVITFAGENGITTKVNGGVVRVGVGIDQTKG-LTPKLTV-GNNNGKGVIVDS--KDGQNT 1169
Db 160 TIIFALAKDLGVKTATVSDTLTIGGGAAGATTTPKVNVSTTDDGLKFAKDAGANGDTT 219
QY 1170 I--TGLSNTLANVNDGAGHALSQGLANDTKTRAASIGDVLNAGFNLIQ-----GNG 1219
Db 220 VHLNGIGSTLTD-TLVGSPATHIDGQDQSTHYTRAASI KDVLNAGWIKGVKAGSTTQGS 278
QY 1220 EAVDFVSTYDTVDFIDGNATKATKYDDTSKTSKVYDVVNDNKTIETVSDKLGKVTFT 1279
Db 279 ENVDVFVHTYDTVEFFLSADTETTTVTDSKENGKRTEKVIKAKTSVKEKDGKLTGKANK 338
QY 1280 LTKTSANGNATKFSAADGDALVKASDI-----ATHLNTLAGDLOF-AKASQA 1326
Db 339 ETNKKVDGANATE-DADEKGLVTAQKVIDAVNKTGWRIKTTDANGONGDFATVASGNTVT 397
```


Thu Sep 13 14:18:21 2001

QY 1327 SSSASYVDA-----DGNKVIYDSTDKKYYQVNDKGOVDKNKEVAKDKLVAQAQTPDGTLL 1380
 Db 398 FASNGTATVNTGDTGITYKDA-----KVGDGKLD-----GDKIAA-----DTTA 440
 QY 1381 AQNVKSVINKEQVNDANKKOGINEDNAFKIGLENAAKOTK---TKNAAVTVGDLNVAQAQ 1437
 Db 441 LTVN-----DGNKANNPKCKVADVASTDEKLVATAKGLVTA--LNSLSW 482
 QY 1438 TPLTFAGDTGT-----TAKKLETLIKGGQDITNKLTDNNIGVAVAGTGTFTVKLARD 1490
 Db 483 TTTAAEADGGTLDGNASDQEVKAGDKVTFKAGK-----NLKVKQEGANFTYSLODA 533
 QY 1491 LTNLSV-----NAGGTGRIDEKGFSEFDANGQA--KANTPVLSANGLDGGKRISNIGA 1542
 Db 534 LTGLTSITLGTGNNKAETINKDGLTTPANGAGANNANTISVTKDGISAGQSVKRVWS 593
 QY 1543 AVDDNDVNFKQFNEVAKTVNNLNQNSGASLPVPVTDANGKPKINGTDGKPKOKAIKAGD 1602
 Db 594 GLKKFGDAN---FDP LTTSSADNLTKONDD-----AYKG-- 623
 QY 1603 GKYHANGVVPVDKDKPITDADLANLAHAHGKPLDAGHGVASLGGNSDAITLTNIKS 1662
 Db 624 -----LTNLDE-----KGTDRQTPVAD----- 641
 QY 1663 TLPDIDPTGNAGAOQSLPSLSAAQOSNAASVKDVLNVGFNLQTNHNOVDVFVAYDT 1722
 Db 642 -----NTA-ATVGDRLGVLWISADKTTGGSTE-----YHDQ---VRNANE 678
 QY 1723 VNFVNGTGADITTSVRSADGTMSTNTVNTALAAATDDGDNVLIKAKDG-----KPYK 1772
 Db 679 VKFSGNGINV-SCKTVNG-RREITPELAKGEVVKVNEFTVKTETNGKETSILVKVGDKYYS 736
 QY 1773 ADDLMPNGLKAGKSADAKTPTGLSLVNPNAKGSTGDAVALNLSKAVFKSKDGTGTTT 1832
 Db 737 KEDI-----DGLTQPKLKDGTVA 756
 QY 1833 TVSSDGISIOGKDNSSITLSKGLNVGKVISNVGKGTGKTDAANVQOLNEVRNLGLGN 1892
 Db 757 AKYQD-----KGGKVS-----VTDNTEAT-----IIN 779
 QY 1893 AGNDNADGNQVNIADIKDPNNGSSNRTVIKAGTVILGGKGNNDTEKATGGVOGVGDKD 1952
 Db 780 KGSYVTVNGOV-----ADAIKSGFELGLADE 806
 QY 1953 GNANGDLSNVVVKTKDGSKALLATYNAAGTNYLTNNPAAEADRIEQQIIRFFHNDG 2012
 Db 807 ADA-----KRAPDDKTKAL-----SAGTTELV-----NAHDK-----VRE----- 836
 QY 2013 NQEPVVQGRNGIDSSASGKHSVAIGFQ-----AKADGAAVAIGRQTOAGNQSI--AIGD 2065
 Db 837 -----ANGLTKVSAATVESTDANGDKVTTTFVKTDELPLTQIYNTDANGKKTIVKVKD 891
 QY 2066 -----NAQATCDQSIAGTGNVTVGKHSAGIDPSPVTKADNSYSGVNNNQIFIDATQT 2117
 Db 892 GQTKWYELNADGTADMTKEVTLGNVDS-----DGKVKVWDND-----KGEVSN 949
 QY 2118 DVFVGNNIIVTESNVALGSNAISAGTHAGTQAKKSDGTAGTTTATAGTGVKGFAGQ 2177
 Db 929 -----GKWYHA-----KADGTADKT----- 949
 QY 2178 TAVG-----AVSVG-----ASGAERRIONVAAGEVSATSDAVNGSOLYKATQOIAN--A 2225
 Db 950 DKVSTDEKHVSLDPNDQSKGKGVYVDNVANGDISATSDAINGSOXYAVAKGVNLAGQ 1009
 QY 2226 TNELDHRLHONENKANAGISSAMAMAMPQAYIPGRSMWTGGTATHNGOGAVAGLSKLS 2285
 Db 1010 VNNLEGKVNKVKRADACTASALAASQLPOATMBPKSNVAIAGSSYOGONGLAIGVSRIS 1069
 QY 2286 DNGQWFKINGSDTQGHVGAAGVAGHF 2314
 Db 1070 DNGKVIIRLSGTTNSQGGTGVAAAGVGQW 1098

RESULT 8
 Q9F3X6 PRELIMINARY: PRT: 1299 AA.
 ID Q9F3X6; AC Q9F3X6; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE MAPA PROTEIN.
 GN MAPA.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OC NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
 RT "Evolutionary origins of the autotransporter proteins";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ277635; CAC14202.1;
 SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;
 Query Match 5.69; Score 653.5; DB 2; Length 1299;
 Best Local Similarity 23.3%; Pred. No. 1.3e-19;
 Matches 402; Conservative 207; Mismatches 554; Indels 563; Gaps 84;
 QY 699 GSKDKGKTOLVIEQVAVASGNDTKNIIRGLSLTPLS---ITNAGGYRTTE-OGNTTTSDEK 754
 Db 27 GKSASKSALV-----NSVSGFSFTLIAASVVLGSGQVNAAEVTCNTGVSGDDK 75
 QY 755 SKAASIGDILNFGNLKNNNSVGFVSTVNTVDFIDGNATTAKVTYDETNTSKYTVDN 814
 Db 76 -----YCYNASSO-----SVICGDATTK--TTDKTD-----N 101
 QY 815 VDEKTIEL---TGDNKGTN-KIGVKVTKTTLTTNANGKATNPFSTTNDALVNAKOIAEHLN 870
 Db 102 KPAKSVVIGFATNDGETNVAIGAK-----SKSKAASIAIGDNKALDNOQAIGON 154
 QY 871 TLA-KEHTTKGTADTALOTFKVKGDKATDDETITVKGDKGTONGKTVNTL-----K 920
 Db 155 ATANDWDISIGROAGAEQI-EVSAEG---RNIAIGDGLAKRGKGVNNNIALGTSAGDR 209
 QY 921 LKENGTLTAV--NKDGTVTFGINTQSLKAGDS-----TTLNKGDSLKNPAS--- 967
 Db 210 LAGTHNVLMTGVYNADEAV-----RSALTAGSTTKEINAKETDNKYITEASNTVALGT 263
 QY 968 -----NEQIOVGADGVKFAKVDKNGSSTGIDGTSRITKDOIQTGANGSLDTPKHLTK 1021
 Db 264 RALATQLAAVAIGQAKAF-----GNQSVAVNGTKASGTATATGSNA-----HAT- 310
 QY 1022 DKLKVEVEITNTGINAGKKITNIQSDIT-----QNSNDAVTGGRVYDLKTELES 1073
 Db 311 -----GSSSIAIGGVNGNDYARTLASDNFTIAMCLSAQATKSDAIAVGR----- 355
 QY 1074 KINSAAKTAQNSLHEFSVADEQGNHFTV-----SNPYSSYTSKTSVDITFAGENGITKV 1129
 Db 356 ----NAKAA--GINTVSGYGNAGVTKTADAQSDPNKLEPA--TDAVFIGNKAGYKSNQ 407
 QY 1130 NKGVRVVGIDQTKGLT-TPKLVIGNNNK-----GVIDSKDQONTITGLSNTLANVTN 1182
 Db 408 NRMQVSLGKDSGEGYVGTENTVIGNSAGKNTKGTNTVAISSRAGON-VEGHIDNEFAA--- 462
 QY 1183 DGAGHALSQGLANDTDKTRAASIGDVLNAGFLNQNGEAVDFVSTYDVFIDGNATTAK 1242
 Db 463 -----LIEAQNIKGS-----DNIAIGK 480
 QY 1243 VTYDDTSKTSKVVVDVNDVKTI-----EVTSDKKLGVKTKTTLTKTSANGNATKFSAAQGD 1298
 Db 481 HAGRSADPNTK-----LNINN-TISLKGESVSLKNFGI-----AQGNKAK---TDG- 522
 QY 1299 ALVKASDIATHLNTLAGDIQTAKGASQASSASVYVDADGNKVIYDSTDKKYYQVNDK--G 1356

Db 523 ----LASIAIGRNAEAVGGTETANIAIGDSASA-----DASGAIVLGTAKAQSITVDGKKYG 575
QY 1357 QVDKNEKAVKDLIVAQA-----QTPD-----GTLAQMNVKSVINKEQVNDANKKOCINE 1405
Db 576 AYSIAVIGTEAKAIAQAAPAGRNENPKDAIAGTKRAEHYASTI-----619
QY 1406 DNAFIKGLENAKDKTKNAVTVG--DLNAVAQTPTTFAGDTGTTAKKILGETLTIKGGQT 1464
Db 620 -----ALGFAGK--SDTKAQAVSIGYNSNAKGQAIAFGSEAKTT--ENAGSSIAF--GTRKA 670
QY 1465 DTNKLTDNIGVYACTDGTFTVKLAKDLTLNLSNVNAGGTRIDEGISFVDANGQAKANTPV 1524
Db 671 QTRASASIAIGMAET--GFD-----GGQALD--GSDAVALGREAKAK---709
QY 1525 LSAINGLIDLGKRISNITGAADVDDNDVNRQFNEVA-----KTVNNLNQNSGASLPF 1577
Db 710 -RONALAFYK-----AVADHKDAVALGAGAEATAAEGTNEATYNEFYSGFAGI---758
QY 1578 VYTDANGKPIGTGPKQKAIKAGDKYTHANANGVDPDKGPTDADKLANLAHGRP 1637
Db 759 -----KPI-----ATVSVGKKDAERTITNVAAGRDKTSDAIN-----792
QY 1638 LDAGHOVASLGSNSDAITLTNKTSLTLPQIDPTPNTGNANAGQAQSLPSLSAAQSNAAV 1697
Db 793 ---GSOLYLALN-----ALGNVGNFL-----VTNVLGGDA-----AIVKEGDEAGT 830
QY 1698 KDVNLVGNFNLQTNHNDVFNKAYDITVNFVNGTG--ADITSVRSADGTMSNITVTNATAATD 1756
Db 831 LPMNSNIG---GTCKGTI-----HDAEAVNNTAKASKTTVKEGD-----NITVEEAAADG 878
QY 1757 DGNVLIAKADGKYFKADLMPNGSLKAKSASDAKTPGTGLSLVNPNAKGSTGDAVALN 1816
Db 879 SRTYTVATKDKVFE-----DSVAVAGTKIDA-----NGLTFVD-----911
QY 1817 NLSKAVFKSKDGTITTTVSDGISIOGKDNSSITLSKDLNGLVGGKVISNVGKG--TKDPT- 1873
Db 912 -----DQGT-----KIDNTPSTSKTGIDAGNKVTNVQNGNIAKDSK 948
QY 1874 DAANVOOL-----NEVRNLLGLGNAGDNADGNQVNIADIKDPNSGSSNRTVIRAKTTL 1929
Db 949 DAVNGQLFAQEGEVKNIG-----GDTYTPETGEYAN-----TNI 985
QY 1930 GKGKGNNDTEKLATGGVGVGVKDGNGDLSNVWVKTKDGSKKALLATYNAAAGQTNILT 1989
Db 986 GGTGASTI-----DEAIKAVNTTAKAA-----1007
QY 1990 NNPAAIDRINEGIRFFHVDNGNQEPVYVQVRNGIDSSASGRHSVAIGFQAKADGEAANA 2049
Db 1008 -----KTEVVQGENIVVTSAPGAN-----1026
QY 2050 IGRQTOAGNOSTAIGDNAONTGOSTAIGTGNVVTCKHSGAIGDPS--TVKADNSVSGN 2107
Db 1027 -----GNTVYVATAKEVTFDKTI-----VGSVVTDKNTDITGLSNKLTGDNFAKNR 1076
QY 2108 --NNQFIDATQTDVFGV--GN-----NITVTSNLSVALGS--NSAISAGHTAQAK--2154
Db 1077 AASEEQLNATQTNLATLLGNAONTNGNVAMTDIGTGKNNINDAIKASRNEVKQCKNMV 1136
QY 2155 ---SDGTAGTTTAGTGTGVKFGAGTAVGAVSGVSGAERRIQNTYAAGEVSATSTDVN 2211
Db 1137 VPTTGANGOTIYEVATAOKVAF--DEVKVGGITIDAT--TNKISGIAKGDISNSTDAVN 1193
QY 2212 GSOLYKATOGCIANA---TNELDHRIHQENKANKANAGISSAMAMASPMQOYIPGRSMVTGGI 2268
Db 1194 GSOLYELQOKIAKSGDNYNLLNRINKVDKDLRAGTAGANAAAGLQPAQYIPGKSMVAVAA 1253
QY 2269 ATHNGQAVAVGLSKLSDNGQWFKINGSADTQGHVGAAGVAGFHF 2314
Db 1254 GTYKGNALALGMSRISDNGKVKIKLTGNTNSRGDFGASTIGAGYQW 1299

RESULT 9

Q9VTK8
ID Q9VTK8; PRELIMINARY; PRT; 2586 AA.
AC Q9VTK8;
DT 01-WAY-2000 (TREMBLrel. 13, Created)
DT 01-NAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-WAR-2001 (TREMBLrel. 16, Last annotation update)
DE CG18331 PROTEIN.
GN CG18331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Chew S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puril V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2193(2000).
DR EMBL; AE003544; AAF50040.2; -
DR FlyBase; FBgn0036181; CG18331.
SQ SEQUENCE 2586 AA; 260194 MW; 8BBB2435A9PEAE5B CRC64;

Query Match 5.0%; Score 588; DB 5; Length 2586;
Best Local Similarity 19.0%; Pred. No. 1.9e-16;
Matches 513; Conservative 369; Mismatches 732; Gaps 105;

QY 19 VAECASKSHSGSS-----SSTAGOVGSS--PVILRTVATLALIVIGATLNGS 64
Db 168 VVEVSQGTNGNGSSSTSSSTSTTTTSSDEGQTTSSSDPVVVEVAQ-----GSSSNGD 218
QY 65 AYQNNKSTAFGTTGNNDNASASNEASIAIGSLAKAHANOAIAIGGSKPDPNROANOKA 124
Db 219 GNSTQSLTITTTTTTSSDGGQSTSSDPVVEVSQ-----GTNGSNSSTQSSSTTTT 271
QY 125 GSHAKGKESIAIGGDV-LAEGDASIAIGSDDLYLDRNSTNSKYPNGLLSTLQIONHTVLRQ 183

Thu Sep 13 14:18:21 2001

Db	1205	---QGTNGDSSSTQSSSTTTT---	SSDEGQTTSSAPVVEV---	TQG-----	1244
Qy	1152	GNNGKGIVIDSKDQNTINGLNTLANVNDGAGHA---	LSOQLANDTKTRA	1202	
Db	1245	SSNG-----	DGNSTQSTTTTTTTTTSSDGVVVEVSGTNGDSSSTQ	1296	
Qy	1203	ASIGDVLNAGNLCGNGEAVDFVS--TYDVTDFIDGNATTA	KAKYIDDTSKTS-----	1252	
Db	1297	SSSTTTTTSSDEGQTTSSAPVVEVTOGSSNGDGNSTQSTTTTTTTTTSSDGGSTTS	1356		
Qy	1253	-----KVYDVNVNOKYTEVTSKGLVKTKTTSANGNATKFS	AAADGD-ALVKAADI	1306	
Db	1357	SDPVVEVSGTNGDSSSTQSS-----	STTTTSSDEGQTTSSSGPVVDISOSSSNV	1409	
Qy	1307	ATHLNTLAGDI-----	-----QAKGASQ-ASSS	1329	
Db	1410	LNDNNYNDIDFRWANPPVLTPLWKSVEQMAITAPLSLRRP	POOQTSDEGQTTSSS	1469	
Qy	1330	ASYVD-----	ADGNKVYDST-----	DKKIYQ	1351
Db	1470	APVVDISOSSSNGDGNSTQSTTTTTTTTTSSDGGESTT	LSDPVVEVSOCTNGDNNSTQ	1529	
Qy	1352	-----VNDKGQVDKNKEVAKDKLVAQAQTPDGT	LQAMNVKSVINKEQVNDANK---	1399	
Db	1530	SSSTTTTTSSDEGQTTSSSDPVVEVVAOQSSNGDGNSTQSTTTTTTTTTSSDGGESTT	1589		
Qy	1400	-----KOGINEDNAPIKLENAAKDPTKNAATV-----	GDLNVA	1436	
Db	1590	SSDPVVEVSGTNGDSSSTQSSSTTTTTTTTTSSDEGQTT	SSSDPVVEVVAOQSSSNGDGNSTQ	1649	
Qy	1437	QTPLTAFAGDTGTTAKKLGETLTIKG-----	GOTDNTKLTNNIGVVAG	1479	
Db	1650	SSSTTTTT-----	SSDGGESTTSSDPVVELRDLRMVMTETLSPRQRLQRLHPMVAN	1707	
Qy	1480	TGFTVKLAKDLNMLSVNA-----	GGTRIDERGISFVDAN	1515	
Db	1708	PPHLTPLWKSVEQTAITAPLSLRRPQORHLQRAKLPL	LLSDPVVEVVAOQSSSNGDGN	1767	
Qy	1516	GOAKANTPVLISANGLDLGGKRISNIGAAVD-----	DNDA-----	VNFK---	1553
Db	1768	STQSTTTTTTTTTSSDGGESTTSSDPVVEVSGTNGDSSS	SSSSSTTTTNEVSLKDNR	1827	
Qy	1554	---QFNEVAKTVNNLNQ--SNSGASLPFVFTDANGK	PNGTDGKPO-----	KAIGK	1600
Db	1828	SPKNRTTKTYSSTIRIPNSGRKLNSSSETSTVT	SSSSSKPQTKYSSSSSKSNNG	1887	
Qy	1601	ADGKY-----	HANANGVPVDKDKPITDADKLANLAHCKPLDAGHQVVASLG--	1649	
Db	1888	GKNKYWKRTKTKSKNNNG-----	SSTIVGEESDLSL-----	DAGVDVTQGNLN	1935
Qy	1650	---GNSDAITLNIKSTLPQIDPTNGNAGAOQSLPS	LSAAQSNAAVSKDVLNVGFNL	1707	
Db	1936	DEGNSSQSTVT---SSLPVVDI-----	SADVONSESLTSTENTTKYSSK-----	SFKV	1981
Qy	1708	QTNHNOVDFVKAYDTVNFVNGTGADITSVRSADGTMSNI	-----TVNTALATDDGNVL	1762	
Db	1982	PKSNGQ--SSISAKTKTVTSTSTPNVKS	SKSKTSNGSKSVKTSSTTTTSSDPC--	2038	
Qy	1763	IKAKDKFYKADDLMPNGSLKACKSASDAKPTGLSLVNP	NPNAAGKSTGDAVALNLSKAV	1822	
Db	2039	-----	QSSSTQIQIPNDIKSLNOVTITSSVSQVGPSPV--	KVTKET	2082
Qy	1823	FKSKDG-----	TTTTTVSSDGISIOGKONSITL--SKOGLNVGKV	ISNVCKGTDTDA	1875
Db	2083	SVSKDGKTRSSSTTTTTTKGSNO---	SGTLTLPVADGLKSSKTKTTTTTKGTLKSD-	2138	
Qy	1876	ANVOOLEVNRLLGLGNAGDNDAGNOVNI-----	ADIKDPNSGSSSNRTVKA	1925	
Db	2139	---ILSLPEVDASIAVNGDESRSASIKDTNLLSK	LDLSLPKDLASLVNNGKSSKSSSTTT	2196	
Qy	1926	GTVLGGKGNNDTEKATGGVQGVGDGNANGOLSNVVK	TOKDGSKKALLATYNAAGOT	1985	
Db	2197	---TSTKGNKYSLSLPEVDASIAVNGDARSASIKDTN	ILSKLDLSLPKDLASLN-----	2250	


```
QY 1986 NYLTNPAAEALDRINEGIRFFHVHNGNQBPVVGRRNGIDSSASGKHVSVAIGFOAKADGE 2045
Db 2251 -----VNGKSSKSSSTTTTSTRGKLSLSL-----PEVD 2282
QY 2046 AAVAIGRQTAGNQSAIGNAQAGDOSTAI-----GTGNVVTGKHSGAIGD 2093
Db 2283 ASIAVN-----GDAKASAIKDTNLSKIDLSLPKDALUNVNGKSSKSST 2330
QY 2094 PSTVKADNSYVGNNOFIDATQDFVGV-----NNITVESVALGNSAI 2142
Db 2331 TTTTSTNG---SKSKILTPKPKDA-GISIDGIGSGTSTKIKITSNKAAPKASSSF 2386
QY 2143 SAGTHA-----GTQAKKSDGTAGTTTATGATGVKGFAGQATVAGVSGASGAERR 2193
Db 2387 KTTTSTSTSSVPKTESKYSWSSSKTSNPRLTLPINAGISVGGGDSGGSWSKLI 2446
QY 2194 IONVAAGEVSATSTDAVNGSOLYKATQGTANATNEIDHRI--HONENKANAGISSAWAMA 2251
Db 2447 KRSTSDETNASDGLSGSIDSRVTPVPRDLRDIRLGRQSGNDAQSN-----2501
QY 2252 SMPQAYIPGRSMVTGGTIATHNG-----QGAVAVGLSKLS-----DNGQWV 2291
Db 2502 SWTQATQGSSESLANGAITANGLSLEGSSGGVATTIPGSGVGTQYPPYWGNGRWV 2560

RESULT 10
Q9RL69 ID Q9RL69 PRELIMINARY; PRT; 2478 AA.
AC Q9RL69;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DE MRP PROTEIN.
GN MRP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=97302526; PubMed=9158773;
RA Wu S., de Lencastre H., Sall A., Tomasz A.;
RT "A phosphoglucomutase-like gene essential for the optimal expression
of methicillin resistance in Staphylococcus aureus: molecular cloning
and DNA sequencing.";
RL Microb. Drug Resist. 2:277-286(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=97431478; PubMed=9286983;
RA Jolly L., Wu S.W., Van Heijenoort J., de Lencastre H.,
Mengin-Lecreulx D., Tomasz A.;
RT "The femR315 gene from Staphylococcus aureus, the interruption of
which results in reduced methicillin resistance, encodes a
phosphoglucomutase.";
RL J. Bacteriol. 179:5321-5325(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=99265121; PubMed=10332717;
RA Wu S., de Lencastre H.;
RT "Mrp-a new auxiliary gene essential for optimal expression of
methicillin resistance in Staphylococcus aureus.";
RL Microb. Drug Resist. 5:9-18(1999).
DR EMBL: Y09927; CAB55329.1.
SQ SEQUENCE 2478 AA; 263029 MW; 6B9859A02D023C74 CRC64;
```

Query Match 4.9%; Score 578; DB 2; Length 2478;
Best Local Similarity 20.8%; Pred. No. 4.7e-16;
Matches 545; Conservative 329; Mismatches 1033; Indels 708; Gaps 120;

```
QY 67 AQNNKIAFTTGNNDNIASSNEASTAIGSLAKAHANQAIAIGSKPDPNRQA-----NQ 122
Db 100 AQPNTQPA-----CQGNADPNNAQAQPGNOA-TPANQA-CQGNNOATPNNAATPANQOTQ 153
QY 123 KASHAKGKRESIAIGGDVLAEE-----GDASI-----AIGSDDLYLDRNSTNSKY 166
Db 154 PANAPAAQAAPVAANAQOTDPNASNTGEGSINTTLTFDDPAISTDENQDPTVTVTYDK 213
QY 167 PNGILLSTLIQ-----HTVLQ-----IRDSNGSOKYRRTAAEGHASTAVGAMAYAKGHFAN 218
Db 214 VNGY--SLINNGKIGFVNSSELRSMDKNNPQNYQ-----AKGNA- 253
QY 219 AFG-----TRSTAGNYS-----LAVGLT---AKAEKGYTIAIGSNAQAINYG 258
Db 254 ALGRVANANDSTDHGNGFISKTVNVKPDSELIINFMTQNTNSKQCATNLVKAOKKNTTEL 313
QY 259 ALALGADT-----RVDLDY---GIALGYGSOILNNNNNNNNKAYPE---GN 299
Db 314 ATVNVAKTGAHLFKVPTDADRLDLQIPONTAVADASRITTNKDKGKYKYSFIDNVLGFS 373
QY 300 GSN1-----KSSKATCNGLFSIGSSITKRKIINVAG-----YEDTDAVNVQOLKAV 346
Db 374 GSHLYKNRDLAPKATNNKEYTINTEIGNNG--NFGASLKADQFKYEVTLPQGVY---V 428
QY 347 ENLAKRQITF-KGDDNGTGVKKLGETLTIKGGETQADKLTDNNNIGVVTDNNT-----399
Db 429 NN--SLTTFPNGNEIDSTVLK-----NMTV-NYDQNAKVTFTSQ-GVTTARGTHTKEVL 479
QY 400 ----GLKVKLAKNLGLSETVSTKNLTASEKVTVSGSNGNTAELQSGGLTFTPTTNASTDKT 455
Db 480 FPKSLKLSYKVNANIDT--PKNIDFNEKLTARTASDVV-----INNAOPEVT 526
QY 456 VYGTDLKFTDNSNTALEDTTRITTKDKIGFSNAGTVDENKPY-----LDKDKLKVGNS 510
Db 527 L-----TADPFSVAVEMNKDALQQVNSQVDSNHYTTASIAEYNKLLQOQADT 573
QY 511 LNNGGLTVNNTIGGSNKQIQVGADG-----IKFADVNVNVNSNAAKFQTTTR 555
Db 574 ILNEDA---NHRVKTANRASQADIDGLVTKLQAALIDNOAAIAELDTKAQEKVTAQOQSKK 630
QY 556 ITEEEIGFADADGKVDKSPYLD-KKQLOVGVYKITKDSGINAGDQKISNVKDATDFTA 614
Db 631 VTODEVAALVTKINNDKNNIAEINKQTTAGVYTFEKDNGI-----AVLEQDV 678
QY 615 VTVKQLKQVQDADGAL-----QSFSIRDEKGOEFTISNLSYNGNTPNTFTITFA 665
Db 679 ITPTVRQAKQDIIQAVTTRKQIKKSNASIQDEK-----DVAN-----DKIGKTIKA 729
QY 666 GENGISINDIAKGVKVGIDPIN---GLTTPKLTVSGDKGKTQLVIEQVAGSNDTKN- 721
Db 730 DIDAAATTAQVEAIKTK-AINDINQTTTATKAAALAEEDFVQVQAQIDQAPLNPDTTNE 788
QY 722 ----IIRGLSPTLPSITNAGVTRTEQNTITSDDEKSKAASIGDILNTGFLKNNSNSV 777
Db 789 EVAEAIERNAAKVSGVKAIEATTTAODLERVNEEKIENITDSTQTKMDA---YNEV 845
QY 778 GFVSTYNTVDFIDGNATTAKVTYDETNOTSKVT-----YDVNVDEKTIETLTDGNGKT 829
Db 846 KOAATARKAQ-----NATVSNATNEEVAEADAADAAQKQGLHDIQVVKSQEVAADTKSV 901
QY 830 -NKI-GVKTTTTLTNTNANGKATNFSTTNDALVNAKDAENLNTLAKEIHTTKGTADTAL 887
Db 902 LDKINAQIQTAKVKPAADTEVENAYNTRKQEQNS-----NAST-TEEQAAAYTELDTKK 955
QY 888 QTFKVKKGATDDETTIVGKDGQTQNGKTVTVTLKKGENGLTVATNKDGTCTVFGINTOSGL 947
Db 956 QEARTNLDAANTNSDVTAKDNS-----IAAINOVQAATTK-----KSDA 995
QY 948 KA--GDSTTLNKDGLSLKPNPASNEOIQVGADGVKFAKV-----DKGNSSTGI 992
Db 996 KAEIAQKASERKTAIEAMNDSTTEEQAAKDKVDQAVVTANADIDNAANDVDNAKTTN 1055
```


Thu Sep 13 14:18:21 2001

QY 993 DGT-SRITKO-----QIGFTGANGSLDITTKPHITKDKLKGVEI-- 1031
 Db 1056 PATIAATPDANVPAAKQAIADKVOAQETAIDGNNGS--TTEKAAAKQOVQTEKTTADA 1114
 QY 1032 -----TNTGINAGCK---KITNIQSGDITO--NSNDAVTIGRVRVDLTKTELESKINSAAK 1080
 Db 1115 AIDAAHTNAEVAEAKKAAIAKIEAIOPATTKDNKAEIA-----TKANERKTAIAQ 1166
 QY 1081 TAQNSLHFEFSVAD-----EOGNHFTVSNPSSYDTSKTSOVIITFAGNGI--TTKV 1129
 Db 1167 TQDITAEELIAANADVNAVTOANSNTEAANSQNDVQAKTT-----GENSDIVQTPV 1220
 QY 1130 NKGVRVIGIDQTKGLTTPKPLTVGNNGKGIIVDSKQONTITGLSNTLANVTVNDGAGHAL 1189
 Db 1221 NK-----KATARNEITAAILNKLQEIQATPDATDEEKQAADAEANTENKANKAQAI 1270
 QY 1190 SQGLAN-DTIDKTRAASIGDVLNAGFNLONGEAVDFVSTYDVFIDGNATTAKVTYDDT 1248
 Db 1271 SAATTNAQVDEAKA-----NAAAI-----NAVTPKVVKKQA 1302
 QY 1249 SKTS-----KVVDVNVNDKTIETVSDKLGKVTTLTKTSANGNATKFSAADGDALVKAS 1304
 Db 1303 AKDEIDQLOQTQNVINNDQNATTEKEAAIQOOLATAVTDAKNNT--AATDDNGVDQAK 1360
 QY 1305 DIATHLNTLAGDIOTAKGASOASSASYVDAD-----GNKVIYDST----- 1345
 Db 1361 DAGKN-----STQSTQPATAVKSNKN--DVDQAVTTQNOAIDNTTGATTEKNAAKDLV 1413
 QY 1346 ---DKKYYQ-----VNDKGQVDKN-----KEVAKDKLVAQAOTPDGTILA 1381
 Db 1414 LKAKEAYQDILNNAQTNDVTOIKDQAVADIOGITADTIKDKVAKDELATKANEQKALIA 1473
 QY 1382 QMNVKSVLKEQVN---DANKKQO-INEDNAFIKLENAAKDTTKNAAVTVGDDLNAV-A 1436
 Db 1474 QTADATTEEEKQANQOQVDAQLTQGNQNIENAQSIDDVNTAKD-----NAIOA 1520
 QY 1437 QPTLTFAGDTGTAKKLGELIT-----IKGQDTNKLNTDNNGV--AGTDGFTVK 1486
 Db 1521 IDPIQASTDVKTNR--AELLTEMONKITEILNNNETNEEKNDIGPVRAAYEEG--- 1574
 QY 1487 LAKDLTNLSNVNAGTRIDEXGISEVDANGQAKANTPVLSANGL--DIGCKRISNIGA-- 1542
 Db 1575 -----LNNIINAAITGD-----VTTAKDTAVQVQQLHANPVKPKAGKELDQAAADK 1622
 QY 1543 --AVDNDVAVNFKQNEVAKTVNNLNQNSGASLPFVVTVDANGCKPINGTDGKPKAKIG 1600
 Db 1623 KTQIEOTPNASQOEINDAKQEVDTLNOAKTNVQD-----SSINEYVDNAVK- 1669
 QY 1601 ADGKYIHANANGVPVQDKCPITTDADKLANLAAGKPLDAGHQVVASLGGNSDAITLNI 1660
 Db 1670 -EGK-----AKINAVKTPSEYK-----KDALAK-----TEDAYNAKVNEDNSNASTSEI 1714
 QY 1661 KSTLPQIDTPTNGANAGQAQSLPSLSAAQOASVSKDVLNVGNFNLQTNHNVQDFVKAY 1720
 Db 1715 -----ABAKQKLAELKQOTADQNVNOATSKDDIEVQI-----HNDLDNINDY 1755
 QY 1721 DTNVFNCTGADITSVRSADTMSNITVNTALAAITDDGNVLILAKADCKFYKADDLMPNG 1780
 Db 1756 -TIPTGKESATDLDLAYADOKKNISADT--NATODEKQOAIKQVDQNVQTALESINNG 1812
 QY 1781 -----SLKAGKASDAKPTGLSLVNPAGKG-----STGDVA----- 1813
 Db 1813 VDNQDVALTOGKAAIDAIOVD--ATVPKPAQAIQEVKAEADTKESIDSDOOLTAEBKTE 1870
 QY 1814 ALNLSKAVFKSKDGT--TTTVSSDGISIQGK---DNSSI--TLKQGLNVGGRVINSV 1866
 Db 1871 ALAMTKITDQAKOGITDAITAEVEKAKAAGLEAFDNIQIDSTEROKAIEELETALDOI 1930
 QY 1867 GKGTKDQDAANVQOLNEVRNLLG--LGNAGNDNADGNQVNIADIKKDPNSGSSNFTVIK 1924
 Db 1931 EAGVNVNADAITEEKEAFNTALEIDILSKATEDISD--QTTNLAELATVKNALBQ-----LK 1984
 QY 1925 AGTVLGGKGNNDTEKLTATGGVGVQVDK-----GNANGDLNWNVVKTKQDKGSKALLATYN 1980

Db 1985 AQRI-----NPEYKKNALAEIRVYNKQIIEIKNADADASAKAEIARTDLGRYFDFRADKL 2039
 QY 1981 AAGQTN-----YLTNNPAEATDRINEOGIRFFHVNDGNQEPVVGRRNGIDSSASGKHVAI 2036
 Db 2040 DKQTNAEVAELQNVITPAIEAIVPQNDP--DANDTN-----NGID----- 2078
 QY 2037 GFQAKADGEAVALIGRQTOAGNOSIAIGDNAQA-----TGQDSIAIGTGN-----VVTG 2085
 Db 2079 -----NNDATANSNANATPENTGQPNVSETTANGKADASPTTPN 2117
 QY 2086 KHSGAIGDPSTVKAD---NSYSVGNNOQTIDA-----TOTDVFVGNNIT--- 2127
 Db 2118 NSDAATGETTATSATDANDKPOANNSSYDASTNSTMDNDVTSRPEVESTNNGTDRP 2177
 QY 2128 VTESNSVALGNSAISAGTHAGTOAKKSDGTAGTGTGTAGATGTVKGAGTAVGAVSVGA 2187
 Db 2178 VTETDNATPAESTTNNNST---TTATNENAPTGTATAPTASTEAASSADSKDNASVND 2234
 QY 2188 SGAERRIQNVAAE-----VSATSTDAVNGSOLYKATOGIANA----- 2225
 Db 2235 SKQNAEVNNSAESQSTNDKVAQPKSENKAKAEKDGSDSTNSQSMVESTETILPSADITEPN 2294
 QY 2226 -----TNELDHRHIONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHN 2272
 Db 2295 VPSNTSKDKEESTTNOQT---AGQKSETNVASNEADKSPSKADTEVSNKPSTSASSEA 2350
 QY 2273 GOGAVAVGLSKLSDNGQWFKINGSADTQGHVGA 2307
 Db 2351 KERKTSNLSQKDDTA-----TADTNDTQKSVGSA 2380

RESULT 11
 Q9LCH2 PRELIMINARY; PRT; 2478 AA.
 ID Q9LCH2
 AC Q9LCH2
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE FMTB.
 GN FMTB.
 OS Staphylococcus aureus.
 OC Bacillus/Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OC NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COL;
 RA Komatsuzawa H.;
 RT "Staphylococcus aureus gene for affecting the methicillin
 resistance."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB025716; BAA93438.1;
 SQ SEQUENCE 2478 AA; 262993 MW; 1C118BEE0DB03B34 CRC64;

Query Match 4.9%; Score 578; DB 2; Length 2478;
 Best Local Similarity 20.9%; Pred. No. 4.7e-16;
 Matches 545; Conservative 334; Mismatches 1033; Indels 698; Gaps 121;

QY 67 AONNSKIAFTGTTGNNDNASASNEASIAIGSLAKAHANQAIAGGSKPDRNOA---NQ 122
 Db 100 AQNTQPA---GQGNQADPNNAQAQPGNQA--TPANQA--GQGNQADPNNAQTPANQQTQ 153
 QY 123 KAGSHAKGESIAIGDVLAE-----GDASI-----AIGSDDLYLDRNSTNSKY 166
 Db 154 PANAPAAQAPAAVANAQODPNASNTGEGSINTTLTDDPAISTDENRQDPTVTVDK 213
 QY 167 PNGLLSTLION-----HTVLQO--IRDSNGSQYRRTAAEGHASTAVCAMAYAGHFAN 218
 Db 214 VNGY--SLINNGKIGFVNSELRRSDFKNNPNQYQ-----AKGNA- 253
 QY 219 AFG-----TRTAEGNYS-----LAVGLT---AKAEKGYTIAIGSNAQAINVG 258

Db 254 ALGRVNDSTDHNGFNISKTWNVPDSELIINFITMOTNSKOGATNLVICKAKNTEL 313
QY 259 ALALGADT-----RVDLDY---GIALGVGSQILNNNNNNNNKAYYPE---GN 299
Db 314 ATVNVAKTGAHLFKVPTDADRLDLOFPDNTAVADASRTITNKDGYKYYSFIDNVLFS 373
QY 300 GSNI-----KSKATGNGLFSGSSTIKRKIINVAG-----YEDTDAVNVQALKAV 346
Db 374 GSHLYVKNRDLAPAKNNKTYTINTEIGNNG--NFGASKADQFKYEVTLPOQVTV---V 428
QY 347 ENLAKROITP-KGDDNGTGVKKLGETLTIKGGETAQDKLTDNNNGVVTNNNT----- 399
Db 429 NN--SLTTTFPNNGEDSVLK-----NMV- NYDQANKVTFYSQ-GVTTARGTHKEVL 479
QY 400 ----GLKVKLAKNLGLSETVSTKNLTASEKVTVSGSNTAELOSGGLTFPTTNNASTDKT 455
Db 480 FPKSLKLSKYVNVANIDT--PRNIDFNEKLTVRTASDVV-----INNAQPEVT 526
QY 456 VYGTDLGLFTDNSNTALEDTTRITKDKIGFSNKAQTVDEKNPY-----LDKDKLKVGNS 510
Db 527 L-----TADPFVSVMYEMNKNLALQOQVNSQVNSHYTTTASIAEYNKLKQOQADT 573
QY 511 LNNGLTVNNTIGSNKQIOVGADG-----IKFADYVNVNVSNAKFTGTR 555
Db 574 ILNEDA---NHVKTANRASQADIDGLVTKLQALIDNQAAIAELDTKAQEKVTAAGQSKK 630
QY 556 ITEEEIGFADADGVKQPSYLD--KKQLQVGVKIFTKDSGINAGDOKI-----SNVKDATD 610
Db 631 VTODEVAALVTKINNDKNAIAEINKOTTAQGVYTEKONGIAVLEQDVITPTVKPRANQD 690
QY 611 DTAVTYKQLKVOQDADGALQSFISIRDEKQBEFTISNLSYNGNTPMTFTITFAGNGI 670
Db 691 IIOQVTRK--QOIKK-----SNASLODEK-----DVAN-----DKICKIETKAIDIAA 734
QY 671 SISNDIAKGVKVGIDPIN---GLTTPKLTVSGDKGKTOLVIEQVASGNDTKN-----I 722
Db 735 TTRNAQVEAIRTK-AINDINGTTPATTAKAAALEEFDEVVQADQAPLNDPTTNEEVAEA 793
QY 723 IRLSPLTLPISITNAGGVRTTEQONTITSDEKSKAASIGDILNTGFLKNKNSSVGFVST 782
Db 794 IERINAAKVSQVKAIEATTTAQDLERVKNEEISKIENITDSTQTKMDA---YNEVKQAAT 850
QY 783 YNTVDFIDGNATKATVYDFETNOTSKVT-----YDVNVDEKTELTGDNGKT-NKI- 832
Db 851 ARKAQ-----NATASNATNEEVAEADAADAAQOGLHIOQVVKSKQEVADTKSKVLDKIN 906
QY 833 GVKTTTLTTNANGKATNFSTTDNALVNAKDIAENLNTLAKETHHTTKGTADTALQTFKV 892
Db 907 AIQOQAKVPAADTEVENAYTRKQEIQNS-----NAST--TEEKQAAYTELDTKKQEAR 960
QY 893 KKDGDATDEITVCKDGTONGKTVNTLKLKGENGLTVATNKGDTVTFGINQSGLKA--G 950
Db 961 NLDAANTNSDVTTAKONS-----IAAINQVQAATTK-----KSDAKAIEA 1000
QY 951 DSTLTKDGLSIRNPASNEQIOVGADGVKAKV-----DKGNSSTGIDGT-S 996
Db 1001 QKASERKTATEAMNDSTPEEQAAKDKVDQAVVTANADIDNAANNDVDNAKTITNEATIA 1060
QY 997 RITKD-----QIGFTGANGSILDTTKPHLTCKDLKVGEVEI----- 1031
Db 1061 AITPDANVKAQAIAADKVOAQETAIDGNNGS--TTEKAAAKQOQVOTERTTTADAIDA 1119
QY 1032 -TNTGINAGGK-----KITNIQSGDITQ-NSNDAVTGGRVYDLTKLETSKINSAKTAQNS 1085
Db 1120 HTNAEVEAAKAAIAKIEAIPATTTKDNKAIEA-----TKANERKTAIAQODIT 1171
QY 1086 LHESFVAD-----EQNHFTVSNPYSSYDTSKSDVITTFAGNGI-----TTKVNKGVV 1134
Db 1172 AEEITAAANADVNAVTOANSINFAANSQNDVDQAKTT-----GENSIDOVTPTVKN--- 1222
QY 1135 RVGIDQTKGLTTPKLTGVNNGKGIIVDSKDGONTITGLSNTLANVTNDGAGHALSOGLA 1194

Db 1223 -----KATARNEITAILNNKLOEIQATPDATDEEKQAAADAEAENTENGKANQAISAATT 1275
QY 1195 N-DTDFTRAASIGDVLNAGFNLOQNGEAVDFVSTYDTVDFIDGNATTAKVTYDDTSKTS- 1252
Db 1276 NAOVDEAKA-----NAEAAI-----NAVTPKVVKQAARDEI 1307
QY 1253 ---KVVYDVNVNDKNTIEVTSKGLGVKVTTLTKTSANGNATKFSAAADGDLVALKASDIATH 1309
Db 1308 DOLQATQTNVINNDONATTEEKAAIIOOLATAVTADAKNNIT--AATDDNGVDQAKDAGKN 1365
QY 1310 LNTLAGDIOIATKAGSQAASSASYVDAD-----GNKVLVYDST-----DK 1347
Db 1366 -----SIQSTQFATAVKSNKN-DVDQAVTQNOAIDNTTGATTEEKNAAKDLVLKAKE 1418
QY 1348 KYTQ-----VNDKQGVDRN-----KEVAKDKLVAQAQTPDGTILAOHNVK 1386
Db 1419 KAYQDILLNAQTNDVTOIKDQAVADIQITADTTIKDVAKDELATRANEQKALIAQTADA 1478
QY 1387 SVINKEQVN--DANKKQG-INEDNAFIKLENAAKOTKTKNAAVTVGDDLNAV-AQTPLT 1441
Db 1479 TTEEKQAOQVDAHLTQGNQNTENAGSIDDVNTAKD-----NAIQAIIDPIQ 1525
QY 1442 FAGDTGTTAKKLGETLT-----IKGGQDITNKLTDNNIGVV--AGTDFTVYKLAKDL 1491
Db 1526 ASDVKTNR--AELLFEMQNKITEILNNETTNEEKGNDIGPVRAAYEEG----- 1574
QY 1492 TNLNSVNAQGTTRIDKIGISFWDANGQAOKANTPVLSANGL--DLGKKRISNIGA----AVD 1545
Db 1575 --LNNINAATTCG-----VTTAKDTAVQVQOOLHANPVKKPAGKKELDQAAADKKTQIE 1627
QY 1546 DNDAVNFQPNNEVAKTVNNLNNSGASLPVFFVTDANGKPIGTGKPKAIKAGDGY 1605
Db 1628 QTPNASQOIEINDAKQEVDTLNOAKTNVQO-----SSTNEYVDNAVK--BGK- 1672
QY 1606 YHANANGVPVDKQKPTDADKLANLAHAKPLDAGHQVVASLGGNSDATLTNIKSTLP 1665
Db 1673 --AKINAVKTFSEYK-----KDALAK-----IEDATNAKVNEADNSASTSEI----- 1714
QY 1666 QIDTPNTGNANAGQAQSLPSLSAAQQAASVKDVLNVGNFNLQTNHNOQDFVKAYDTVNF 1725
Db 1715 -----APEAKQLAELAQATADQNVQATSKDDIEVQI-----HNDLDNINDY-TIPT 1759
QY 1726 VNGTGADITSVRSADGTMSNITVNTALAAATDDGNGVLIKAKDGKFFYKADDLMPNG----- 1780
Db 1760 GKRESATDLYAYADQKKNISADT--NATQDEKQQAIKQVDQNVQTALESINNVDNGD 1817
QY 1781 ---SLKAGKSASDAKPTGSLVNPKNAGK-----STGDVAV-----ALNNL 1818
Db 1818 VDALTOGKRAIDAIVQD--ATVKPKANQAEVKAEDTKESIDQSDQLTAEKTEALAMI 1875
QY 1819 SKAVFKSKDGT--TTTVSSDGSISIOGK---DNSSI--TLSKDLGNVGGKVISNVGKGT 1871
Db 1876 KQITDQAKQGITDATTAEVEKAKAQGLEAFDNIQIDSTEKQKAIELEHTALDQIEAGVN 1935
QY 1872 DTAANVQQLNEVRNLLG--LGNAGNDAGNQNVIADIKKDPNSGSSSNRTVTKAGTVL 1929
Db 1936 VNADATTEEKAEFTNALEDILSKATEDISD--QTTNAEIAATVKNSALEQ-----LKAORI- 1988
QY 1930 GKGNNDETEKLGVOVGVDKD---GNANGDLSNVVVKTKDQSKKALLATYNAAGOT 1985
Db 1989 ----NPEVKNNALAEIREVNVNKQIEIKNADADASAKAIEARTDLGRYDFRFDKLDKTQT 2044
QY 1986 N-----YLTNNPABAIDRINBQGRIFFFHVNDGNQPNVQVQNGRIDSSAGKHSVAIGFOAK 2041
Db 2045 NAEVAELQNVTPAIEAIVPQNDP--DANDTN-----NGID----- 2078
QY 2042 ADGEAAVAIGRQTAGNQSTAGDNAAQ-----TGDSQIAIGTGN-----VVTGKHSGA 2090
Db 2079 -----NNDATANSANATPENTQPNVSETTANGKADASPTTPNNSDAA 2122
QY 2091 IGPSTVKAD--NSYSVGNNOQFIDA-----TQTDVFGVGNNT---VTESN 2132
Db 2123 TGETTATSATDDANDKFPQANNSSVDASTNSPTMDNDVTSKPEVESTNNGTTDKPKVETTD 2182

Thu Sep 13 14:18:21 2001

Qy	2133	SVALGNSAISAGTHAGTQAKKSDGTAGTTTATAGATGTVKGFAGQTAAGVAVSGASGAER	2192	Qy	399	TGLVKLAKNLGSETVSTKNLTASEKVTGVS--GNNTAELOS-----	439
Db	2183	NATPAESTNNST--TTATNEAPGTSTATAPTASTEAASADSADKONASVNDKQNA	2239	Db	286	TILK---ATNINLKHNTSTLNLD-DNIIIVIGNIKGNKKDILNFKVHGHTNLDNEMIIPAP	341
Qy	2193	RIONVAAGE-----VSATSDAVNGSOLYKATOGIANA-----	2225	Qy	440	---GGTFTPTTNASTDKTVYGTGDLKFTDSDNTALEDTTRITKDKIGFSNK---AGTV	492
Db	2240	EYNSAESQSTNDKVAOPKSENKAKAEKDSGSTNQSWESTETLPSADITPEPNVPSMT	2299	Db	342	QKTHGTLPF--KGNATLNGNINNLILKFSGGHGTGLNLOGNTKVDNLVFPADSVLDSGTI	399
Qy	2226	-----TNELDHRIHQENENKANAGISSAMAMAPQAYIPGRSMVTGGIATHNGQAV	2277	Qy	493	DENKPYLDKDLKLVGNSTLANGLTVN--NTI-----GGSNKQIOVGADGIFKADVNVV	545
Db	2300	SKDKESTNQTD---AGQLKSEINVASNEADKSPSKADTEVSNKPSTSSASSEAKERMT	2355	Db	400	SVN-GLDTCVTFNNNSVNGGTLIIINAKNTISAKLLNATKAKIQIN-----ANLTMNH	452
Qy	2278	AVGLSKLSDNGOMVFKINGSADTQGHVCAA	2307	Qy	546	SNAARFGTTRITEEIEGFADA-DGKVDKSKPYLDKQLOVGGVKITKDSG-----INAG	598
Db	2356	STNLSQKDDTA-----TADTNDTOPSVGSA	2380	Db	453	PSAGDISIRIADNTIYTIIDAKNGVN-----LLNNAKIIFEGADSMALLINTG	502
RESULT	12			Qy	599	DOKISNVKDATDDTDAVTKOLKOVQODADGALQSFSDIRDEKGOEFTISNLYSNGNTPNT	658
Q92D91				Db	503	-----VTADRTFTIYNLNSQNDYEYIVKIEAK-----	532
ID	Q92D91	PRELIMINARY;	PRT; 2340 AA.	Qy	659	FETITFAGENG-ISISND-----IAKGVKVKVGDIDPINGLITPKLTVGSKDKGTQLV	709
AC	Q92D91;			Db	533	-KVITIANOSGPTYIGQDNTHRLKELIVEGAGDIID--DTIFTKLJSINSTGOITFNRT	589
DT	01-MAY-1999 (TremBLrel. 10, Created)			Qy	710	IEQVAGSNDTKN-----IIRGLSPTLPSITNAGGVRTTEQGN---TITSDKSKAASIG	761
DT	01-MAY-1999 (TremBLrel. 10, Last sequence update)			Db	590	LDLGAGGNIAFAFGKHTLVVNGVGTGSIITSENNOGILITNSGNITGVICTNELGLKLVNIG	649
DT	01-MAY-2000 (TremBLrel. 13, Last annotation update)			Qy	762	-DILNTGFNLKNNNSVGFVSTYNTVDFIDGNATTAKVTYDETNOTSKVTYDVNVNDEKTI	820
DE	CELL SURFACE ANTIGEN (SCA3).			Db	650	ADPYTCSANV---FASVALTNPSVLLADGVTLTGEVT---THNNTKGVLSLG-----	697
GN	RP451.			Qy	821	ELTGDNGKTNKIGVKVTTTTLTNTNANGKATNFS-----TTDNDALVNAK	863
OS	Rickettsia prowazekii.			Db	698	--TGSN-ITGOIGTNSAALEKINIGAGASNIDSNIVAGSTVLTDQOTSELTANDVVASN	754
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			Qy	864	DIAENLNLAKETHTKGTADTALQTFVKYKKGATDDEITTVGKDGTQN-GKTVNTLKLK	922
OC	Rickettsiaceae; Rickettsiae; Rickettsia.			Db	755	IITAGNNSKLIPTGNGGI-----TCNIGANGAALQEVVF---NGTTNIGGTANS----	802
OC	NCBI_TaxID=782;			Qy	923	GENGLTVATNKDGVTEGINTQSGLKAGDSTTLNKDG-----LSIKNPASNEQIOVGADG	977
RN	SEQUENCE FROM N.A.			Db	803	--QNFTVAHSAANVITGLTT-GALKYKDTGTIIAHGGLVGDIDFNKAG-----	849
RP	STRAIN=MADRID E;			Qy	978	VKFAKVDKGNSTGIDGTSRITKDOIIGFTGANGSLDITFKPHLTKDKLKYGEVEIT-NTGI	1036
RC	MEDLINE=99039499; PubMed=9823893;			Db	850	-KFILGD---GAMIDGSVLCN-----GGVAGTLDF-----IGDNVTQNIGA	887
RX	Andersson S.G.E., Zomorodipour A., Andersson J.O.,			Qy	1037	-NAGGKKTINTQSGDITON---SNDN-----VTGGRVYDLKTELESKINSAAKTAQNSL	1086
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,			Db	888	DNANSISTINIO-GDNTKNVTIANDIFVDNIHFTNGGILQL-----	927
RA	Eriksson A.-S., Winkler H.H., Kurland C.G.;			Qy	1087	HEFSVADEQGNHFTVSNPYSSDYTSKTSDDVTFAGENGITTKVNVKGVVRVGDQTKGLTT	1146
RT	"The genome sequence of Rickettsia prowazekii and the origin of			Db	928	-----GGNLTJHN-----IDF-GANGGTLFEN-----	948
RT	mitochondria."			Qy	1147	PKLTVGN--NKGIVIDSKDQONTITGLSNTLANVTNDGAGHALSOGLANDTDTKT-RAA	1203
RL	Nature 396:133-140(1998).			Db	949	-----GNNTYNLNAIIV---NCQN---GILNAFTNLK-----ASDDTICTVKII	986
RL	EMBL: AJ235271; CAA14908.1;			Qy	1204	SIGDV-LNAGFNLOQNGEAVDFVSYVD-TVDFIDGNA-TTAKVTYDDTSTKSK-----	1253
DR	SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;			Db	987	NIGQICTPQNFITQVNNKNTLVSSVNSINFECDANSOLLISAPVDQTIKFINNLNETGG	1046
SQ				Qy	1254	--VYVDVNDNKTIETVSDKILGVKTTTLTKTSANG-----NATKFSAA	1295
Query Match	4.9%; Score 571.5; DB 2; Length 2340;			Db	1047	GIITLDSNGNLLISGNGIKLGSKNELSLNLIKVKVTVDLDLDIONHOLNNGALF	1106
Best Local Similarity	21.7%; Pred. No. 8.2e-16;			Qy	1296	DGDALVKASDIATHLNTLAG-----DIOTA-----KGASQAS--	1327
Matches 554; Conservative 295; Mismatches 935; Indels 773; Gaps 133;				Db	1107	DDQSITSAKIKNINIGTVAGGATVTLDAINDNFDLNTSGMVKHQDSILELKNSSNTNDH	1166
Qy	59	ATLNGSAYQNNSKTAFGT-TGNNDNASASNEASTAIGSLAKAHANAIAIGGSKPDPRN	117	Qy	1328	--SASVYDADGKN--VIYDSTDKKYYQVNDKGVQDKNEKAVAKD--KLVAQAQOTPDGTLA	1381
Db	14	ASLUGTLFTNSN---ATGTIIPNGSVSLNTDAGLVGGVFNNGDIQ-IVNGGREIKISA	69	Db			
Qy	118	QAANOKAGSHAKGESIAIGDVLABGDASIAIG---SDDLILDRNSTNSKYPNGLLIST	173	Qy			
Db	70	DKANAIIIGGINTLKLPELDFEG---VEVSNVSGIPLNAGEDL-----NINF--GPLKF	117	Db			
Qy	174	LIONHTVLROI RDSNGSQYRRRTAABGHASTAVGAMAYAKGHAFANAFTRSTAGBNYSLA	233	Qy			
Db	118	ISNVTSI-----ITGVGKTFESNIDFA---GKNATLQINKDL-	152	Db			
Qy	234	VGLFAKAEKGYTIAIGSNAQAINYGALALGADTRVDLDYGA---LGYGSQILN--NNNN	289	Qy			
Db	153	-NITTKID-----NTVAGNNGSTTFEGS-----GILSNHIGYTNLSLLGINVNG	195	Qy			
Qy	290	NKAYVPGGNSNKKSKATGNGLFSIGSSTIKRKIIINVAGYEDFDVAVNAQLKAVENL	349	Qy			
Db	196	EAKIYAPANNITNAKN-----INLTHNSILTL	225	Qy			
Qy	350	AKRQI-TFKGDONGT-----GVKKILGETLTIKGETQADKL-TDNNNIGVVYTDNN	398	Qy			
Db	226	CDGNITTLKGINNTTETDGOGILNLAYDLGSSSIITGDIIGNSLDITINVILGSAFENS	285	Qy			

Thu Sep 13 14:18:21 2001

```

1847 SSTITLSDGL-----NVGGKVISNVGKTKDQTDAAVQOOLNEVRNLLGLGNAGNDNADG 1900
1281 ANVLFKDAVOLQTGNTGIGGFLDFNAKNT-----VTLN----- 1314
1901 NQVNIADIKDPNKGSSSRRTVIKAGTIVLGGKGNNDTEKLA---TGGVOGVKDGKGNAN- 1956
1315 NNVNAGTVK--NIGTNGTLLI-----VLGASNLNRVNGIAMLKVGAGNVTTIAGGNVKI 1368
1957 GDLSNVVVKTKQDKGSKALLIATYNAAGQTNLTNNPAPAIIDRINEQGRFFHFVNDGNOEP 2016
1369 GEIOGTGINT-----LTLPAHFKLTG-----SINKTGGALKLNFMMGSSVSG 1411
2017 VVQCRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNQSTAIAGNQAATGDSIA 2076
1412 VV-----GTAANSVGDITTAGATSPASSVNAKGTATLGGT 1446
2077 IGTGNVTVKHSIGAIDPSTVKADNSYSVGNNOQFIDATQTDVFGVGNNTVTESSVAL 2136
1447 TSFANTFT--NTGAV-----TLAKGSITSFAKN---VTATSF---VANSATINFGNSLAF 1493
2137 GNSAISAGTHAGTQAKKSDGTAGTTTTAGA-----TGTVK-----GFAGQTAVGA 2182
1494 NSNIITGS-----GTTLTLCANQVYTGTSFTDTLTLNTTFDCAKSGS 1537
2183 VSVGASGAERRIQNVA--AGEVSATSTDAVNGSOLYKAT 2219
1538 NILIKSGSTLDLSGVSNLALVVVTAINFDMNISPDKYT 1576
RESULT 14
Q55582 PRELIMINARY; PRT: 3029 AA.
AC Q55582;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 308.8 KDA PROTEIN.
GN SLR0364.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RC STRAIN=PCC6803;
RA Tabata S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
[3]
SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D63999; BAA10087.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3029 AA; 308789 MW; BEE0BB57E822A4FB CRC64;

```

```

419 VTF-----TGDSTVTGNTGNATLATISVGAGKATL-----GG 451
858 ALVNAKDIAENLTLAKRIHTTKGTADTALOTPKVKKDGATDDETITVKGDTGQNGKTVN 917
452 ALIKAT-----TTKLTDNASAVTF-----TNPVVVTGAIQNTGNA-----486
918 TLKLKGENGLVATNKKDGTVTGNTGSLKAGDSTTLNKGSLIKNPASNEIOIQVADG 977
487-----NNGIVTF-----TGDSTVTG-----NIGTNALATISVGAGK 518
978 VKF-AKVDKGNSGTGIDGTSRT-KDQIGTFGANGSLDTPKPLTKDKLKVGEVETINTG 1035
519 ATLGGAIIKATTTKLTDNASAVTF-----TNPVVVTGAIQNTGNA-----IDNTG 556
1036 INAGGKIIYNIOSGDTONSNDVATGGRVYDLKTELESKINSAAKTAONSLEHFSVADEQ 1095
557-----NANNGIVTFGDSVTG-----NIGTN--NATALISVG-----587
1096 GNHFTVSNPYSSYDTSKTS-----VITFAGENGITTKVNGVVRVGIQDTKGLTTPKLTVG 1152
588 AGKATLGGAIKATTTKLTDNASAVTF-----TNPVVVTGAIQNTGNA-----G 628
1153 N-NNGKGIIVDSKQNTITG-LSNTLANVT-NDGAGHALSGL-----ANDTDKTPRAISG 1206
629 NANN-----IVTFGDSVTGNTGNATLNVGAGIATLEGAIKATTTKLTNAASVL 684
1207 DVLNAGENLOGNBVDFSVTYDVFIDGNATTAKVYDDTSKTSKVYDVNDVN-----1262
685 TLTNVNAVLTG-----AIDNTTGVNDVNGVLNGLALSQVT-----GNIGTNAL 728
1263 KTIETSDK-KLG---VKTTTLTKTSANGNATKFS-----AAD---GDALVKASDIATH 1309
729 ATISVAGKATLGGAVIKATT-TKLTDNASAVTFNPNVVTGAIQNTGNA-----NGIATF 784
1310 L--NTLAGDIQAK-----CASQASSASYVDADKNVIYDSTDKKYYQ-VNDRGQVD 1359
785 TGDSTVTGDIQNTALATISVGAGKATLGGAIKATTTKLTDNASAVTFNPNVVTGAIQ 844
1360 KKEVAKDLVAQA--QTPDGTFLAQMKVYSVINKEQVNDANKKGINEDNAFKGLE---1414
845 -NTGNANGIVTFGDSVTGNTGNATLNTS---VGAGKATLG-----GAIKATTKL 896
1415 --NAADTKTKNAV-----VGDNLNAVATPLTPAGDTGTAKKLGLETTLIKGGQDTN 1467
897 TYNASAVTFNNVNVVTAIDNTGNAN---NGIVTFGDS-----TGTGNTGN 942
1468 KLTDNNGIVAGTGDGTVKLAKDLTLNLSVNAVAGTTRIDEKGISFVDANGAKANTPVL 1527
943 ALATVNG--AGV-----TLQAGGS-----LDANNIDFGARSTLEF 976
1528 NG-LDLGGKRLSNI--CAAVDDNDA---VNFQFNEVAKTVNNLNQSNQSGASLPFV---1578
977 NGPLDGGGKALPYFKAIAANGNAILNVTKSLTASHLTIGTV-AETNIGAGNI-FAIDA 1035
1579 ---VTDANGRPINGDKPKQAIKAGDGYKHANANGVVDKGRPIPDADKLANLAH 1634
1036 SAGDVTILNAQDIH-----FRAIDLSALVLSNLTGCVN-----NILLAADLVAP 1079
1635 GKPLDAGHOVVASLGGNSDAITLTNIKSTLPQIDTPNTGNANQAQSLPSLSAAQSSNA 1694
1080 G--VDEG--TVVFDG-----VNLGINSNVAGAA-----1106
1695 ASYKDVNLGNFLQNLNHNQV---DFVKAYDTVNFVNGTGDITSVRSADGTMSTNITVNTA 1751
1107 -NIGDVGGNKFNLTLYNAVTITDDVNLLEGQNVLINNADFTS--STAFNAGTIQINDA 1163
1752 LAATD-DDGNVLKAKDGKFKYKAD--LMPNGS-----LKAGKASDAKPTTGLSLVNP- 1802
1164 TYTIDANNGLNLTIPAGNIQFAHADAQLILQNSGNDRTITILGANIDPDNDDEGIVILNSV 1223
1803 NAGK-----GST-GDAVALNLSKAVFKSD--GTTTTTVSSDGIISIQ-----GKDN 1846
1224 TAGKKLTLAGGKTFGGA---HKLQDIVFKGEGDGTAGTTFNTTIVLIDITGQLELGATT 1280

```


Query Match	4.8%	Score 560;	DB 2;	Length 3029;
Best Local Similarity	21.1%	Pred. No. 3.5e-15;		
Matches 562;	Conservative 320;	Mismatches 985;	Indels 794;	Gaps
QY	56	VIGATLNGSAYAONN-----SKIAFGTTGNDNNDNASNEASIAIGSLAKAHANO	104	
DB	432	LIDANATGGGYFETQEPFQDTLGNPLSLNQFALTGIDVGNSTTEREFYEV-----GYSNY	488	
QY	105	AIAIGSKPDPNQAONQKAGS-----HAKGKESIAIGGDLAEGDASIAIGSDDLYLORN	160	
DB	489	QIAVVG---PRLNVSESTOGDKTKLRLVDRGRPSVLTG-----ISFDD-----	527	
QY	161	STNSKYPNGLLSTLIQNHVTLROI RDSNGSQKYRRTAEGHASTAVGAMAYAKGHFANAF	220	
DB	528	-TASFIANEL--TPVSSLTVRAGVIWNNASNDVRQ-----FSFALGA-----PL	568	
QY	221	GTSTAEKNYSLAVGLTAKAERGYTTAIGSNAQAQINYGALALGADTRDLDYGIALGYG-	279	
DB	569	GTFNSPQDTDTIAPIVTSILRNPTSEI--TNADSLUTFQVTFSENQVNVNDVDFILSGTGV	627	
QY	280	-----SQILNNNN-----NNNNKAYPEVGENSNIKSSKATGNGLFSIGSSTIKRK	324	
DB	628	SGATISNVSGSGSVYTVVTGVIANNNGTVNLDFASQNI RD--TANNALSNWTTPTTDEQ	684	
QY	325	I-----INVGAGYEDTDVAWVAQLKAVENLAKROIITFKGDDNGPGVKKKIGETLT	374	
DB	685	YTLONTAPAAISITLADNITADDOIINIAE-----SGQAIPITGT---VGGEFNVGDVT	734	
QY	375	I-----KGG-----ETQADKLT DNN-----NIGVVT DNNITGLKV	403	
DB	735	LTWNDKTFTGAVGAGGLFSINPGSDLIVDADLTIAASTATTDAAGNLGSATDNQ-----	790	
QY	404	KLAKNLSGLETVSTKTNLTASEKV-----TVGSGNN-----TAELOGGLTFPT	447	
DB	791	YTVDTTAPIIIVNDVTDNIINAESGQAIPITCTVGGEFNVGDTVTLTVNGKPFYGT	850	
QY	448	TNASTDKT--VYGTDLKLF TDSNTALEDTTRITDKIG-----FSNKAGTVDENKPYLD	500	
DB	851	VDANGDFSIDLVGDLVNGSDLTIAASVATT---DAAGNPGSASDNQTYTVDTTAPT	906	
QY	501	KDLKV-----GNSTLNGLT VNNYTTIGGSNKQIQVGADGKLFADVNVN	544	
DB	907	INAIADVDDIINAVEAGSPVAVSGTTTGVEDGOVTVTIDGNTYATFVTGNATWF---NIP	963	
QY	545	VSNAAKFTGTRITEEIG-FADADGK-----VDKKSPVLDKQLQVGGVKITKDS	593	
DB	964	VADIANF---EATEEVVAVPSDLAGNPATPATRNITVDIVAP-----AVTIDISDDT	1011	
QY	594	GINAGOOKISN-----VKDATDPTDAVYTKQLKQVQDADGALQSFSSIRDEKQGEFTISNL	649	
DB	1014	GAQAND-FITNDDTLVFNGETAEADSTVVSVDGIE-----IGTV	1051	
QY	650	YNGNTPTNFE---TTTFAGENCISISNDIAKVKVKGIDPINGLTTPKLTVGSKDKGT	706	
DB	1052	TANGAGEWTLDTYFTGLLADGDVLSVT-----ATNPITGNSATATQTITVDDTAPT	1103	
QY	707	QLVIEQVAGSNDTKNIIRGLSPITSLNAGGVRTTEQGN--TITSDEDKSAAASIGDIL	764	
DB	1102	-VTINAIIV-DDIINAVEAGSPVAVSGTTTG-----VEDGOVTVTIDGNTYATFVTGNAN	1151	
QY	765	NTGFNLKNNNSVGFSTYNTVDIFDGN---ATTAKVYTYDETNOTSKVTVYVNVDE---	817	
DB	1156	TFNIPVADIANFEATEEVVATVSDLAGNPATPATRNITVDTTAPTVTIN-AIAVDDIINA	1211	
QY	818	----KTIELTG-----DNGKTNKIGVKYKTTTLTTNANGKA-----TNFSTDDNDA	858	
DB	1215	VEAGSPVAVSGTTTGVEDGOVTVTIDGNTYAT-VTGNAWTFNIPVADIANFEATE-EV	1277	
QY	859	LVNAKDIAEN-----LNTLAKEIHHTKGTADTALQTKFKVKDGTADDETITVGKD	908	
DB	1273	VATVSDLAGNPATPATRNITVDVDAVPAVTDISDSDTGAQA-----NDFITNDDTLFV--N	1320	
QY	909	GTONGKTVNPLKILKNGENGLTVATNKDGTVTFGINTQSGLKAGDST-----TLNKDG---	959	


```
Db 2278 VVDITAPVTINAIIVDDIINAVEAGSPVAVSGTTT-----GV-----EDGQVVTVTIDG 2327
Qy 1856 LNVGKVISNVCK-GTKDTPDAANVQOLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNS 1914
Db 2328 NTYTATVTGNATWTFNIPVADIANFEATEEV-----VATVSDLAGNPAT 2370
Qy 1915 GSSSRVT-----IKAGTVLGGKGNNDTEKLTGGVQVG-----VD 1950
Db 2371 PATRNITVDTTAPVTINAIIVDDIINAVEAGSPVAVSG-----TTTGVDEQVVTVT 2423
Qy 1951 KDN-----ANGDSNVVVKTKQDQKSKALLATYNAAGOTNLTN 1990
Db 2424 IDGNTYATVTGNATWTFNIPVADIANFEATEEVVATVSD-----LAGNPATPATRNITV 2477
Qy 1991 N---PAEAIDRINEQ---GIRFFHND-----GNQE---PVVOGRNID-----SSASG 2030
Db 2478 DTVAPAVTIDSDTGAQANDFITNDTLVFNFGTAEADSTVVVSLDGIETGTGTANGAG 2537
Qy 2031 KHSVAIGFOAKADGEAAVAGRQTOAGN-----OSIAIGDNAQATGDQSIAT----- 2077
Db 2538 EWTLDYTGTLADGYELSVTATNTGNSATATQTIIVDDTAPVTINAIIVDDIINAVE 2597
Qy 2078 -GTGNVTVKHSKA-IGDPSVVKAD-NSYS---VGNNOF-----IDATQ 2116
Db 2598 AGSPVAVSGTTTGVDEQVVTVTIDGNTYATVTGNATWTFNIPVADIANFEATEEVVATV 2657
Qy 2117 TDVFG-----VGNNTVTVTES-----NSVALGNSAISAGTHAGTQAKKSD 2156
Db 2658 SDLAGNPATPATRNITVDTTAPVTINAIIVDDIINAVEAGSPVAVS-GTTTGV-----D 2712
Qy 2157 GTAGTGTAG--ATGTVKGFAGOTAVGAVSGAERRIQNAAGEVSATSTDAVNSQ 2214
Db 2713 GOVTVTIDGNTYATVTGNATWTFNIPVADIA-----NFEATEEVVATVSD-LAGNP 2763
Qy 2215 LYKATQGI-----ANATNELD 2230
Db 2764 ATPATRNITVDTVAPAVNELD 2784

RESULT 15
Q9L800 PRELIMINARY; PRT: 2747 AA.
AC Q9L800;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE RTX PROTEIN.
GN ASX.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=645;
RN [1]
RP SEQUENCE FROM N.A.
RC Braun M., Frey J., Kuhnert P.;
RA "280 kDa RTX protein of Aeromonas.";
RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
RL EMBL: AF218037; AAF27914.1; -.
DR EMBL: AF218037; AAF27914.1; -.
DR InterPro: IPR001343; -.
DR InterPro: IPR002035; -.
DR InterPro: IPR002048; -.
DR Pfam: PF00353; hemolysinCabin; 3.
DR PRINTS: PR00313; CABDNCRPT.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 4.
SQ SEQUENCE 2747 AA; 280202 MW; 208FE380E44A5F37 CRC64;
```

Query Match 4.8%; Score 558.5; DB 2; Length 2747;
Best Local Similarity 20.3%; Pred. No. 3.6e-15;
Matches 571; Conservative 350; Mismatches 983; Indels 907; Gaps 140;

```
Qy 7 VIFNKATGTFMAYAECAKSHSGSSSTAGQVSSPVIRLTVATLAILVIGATLNGSAY 66
Db 12 VVVTQLKNIYLVAA-----DGSKQLAEGDI-----LPRDAVL-ITPEGASFNG--- 55
Qy 67 AQNSKIATGTTGNNDNASNEA-----STAIGSLAKAHANOAIATIGSKPKDPNQA 119
Db 56 --GNQFTTLPSTNEQAEDETSQLAQNODVATGTPDEISALQQAILGGADPTQAFEA 113
Qy 120 ANQKAGSHAKGKESI-----AIGGDVLAEGDASI-----AIG 151
Db 114 A-RAGGAPAGGNGNIGVAGSNGGFFVTDRTGDTTIAAAGFDATANQTDAGVTGDAAG 172
Qy 152 SDDLY-----LDRNSTSKYPNGLST-LIQNTVLRQIRDSNGSKYRRTAEAGH 201
Db 173 EDDPFDLVPPVITVAPDNTDPTLTGTTPDAPAGSTVTLVLLTDANGNQOQLTTTVNPD 232
Qy 202 ASTAVGAMAYAKGHFANAFGRSTAEAGNSIAGVLTAKAEKGYTI-----AIGSNAQAIN 257
Db 233 GTFSVDVVT-----PLADGSYTVTANVTPDAGNTATATDGGSDVSTAPSITV 279
Qy 258 GALALGADTRVLDYGLALGYSQI-----LNNNNNNKAYVPEGNGNIKSS----- 306
Db 280 DAPDNTDPTTITGTTDAAGSSTVTLVVTDANGNQOQLSATVQPNCTYSTDVTSPDPG 339
Qy 307 -----KATNGLESIGSSSTIKRIINVGAGYEDTDVAVNAOLKAVENLA 350
Db 340 DYDVTSVTDPAAGTATDGGSDVDVTAPTL---AITLDANITDDDDVINAEE-----A 389
Qy 351 KRQITFGDDNGTCVKKLGELT-----IKGGETQAD----- 383
Db 390 GQQIPVSGTVSG---EPKAGDTVTLTVNGKPTGPDADGRFTILVAGSLAADRTDID 446
Qy 384 -----KLTDNNNIGVVTDNNNTGLKVLKLNLSLETSTVTKNLITASEKVTVGS 430
Db 447 ASVTSTDAAGNSATATSEGYVDVTPAVDLAITLDANITDDDDVINAEE---AQOQIPV-S 503
Qy 431 GNNTAELOSGLTETPTN-----ASTDKT-----VYGTDLG 462
Db 504 GTVTGEFKAGD-TVTLTVNGKPTGPDADGRFTILVAGSLAADRTDIDASTVSDA- 561
Qy 463 KFTDNTALE-----DTT-----RITKDKIGFSNKA-----GTVDENKPY 498
Db 562 --AGNSATATDSEGYVDVTPAVDLAITLDANITDDDDVINAEEAQOQIPVSGTV----- 613
Qy 499 LDKKLKLVGNSTLNGGLTVNN-----TIGSNKQIQVAGDGIRFADVNV 543
Db 614 --SGEFKAGDVT---LTVNGKPTGPDADGRFTILVAGSD---LAUSDRTIDASV 663
Qy 544 NVSNAKEFTTTRITEEIGFADAGKVDKSPYL--DKQLQVGVKITKD----- 592
Db 664 TSTDAAGNSTTKSDTENY-----TVDTVAPDLGIDLPIAVGDNVNVNQAEDGKTPV 716
Qy 593 --SGINAGDQKISNVKDATDQDVAITYKQKQVQDADGALQ-SFSIRDEK-----GQ 642
Db 717 TILSGTVNGD---ANVGDTVTLTLGDSKLTQVTVLNGQLGFSTSTADKLGVSSVTA 773
Qy 643 EFTISNLVSNGTPTNFTETFA-----GENGISISNDIAKGVKVGVID 686
Db 774 EISVTDAAAGNSTTKSDTENYVATTAPALGIDLDPVVGDNVNVNQAEE--ADGKTPVTL 831
Qy 687 -PINGLT---TPKLTVGSDDKQTLVIEQVASGN-----DTKNIINGLSPTLP- 731
Db 832 GTVNGDANVGDIIVTLTLG---DG-SKLTQVTVLNGQLGFSTSTADKLGVSSVTA 887
Qy 732 SITNAGGVVTTTQGN-----TITSDEKSKAAASIGDILNTGFLNKNNSNVG 778
Db 888 SVTDAAGNSTTKSDTENYVDNVAAPATVTTIIDDAN-----NDQLLTKEI--GSDQIQ 940
Qy 779 FVSTVNTVDFIDGNATTAKTVDNETNOTSKVTYDVNVDEKTIETL-----GNGKTN 830
Db 941 VRADVNHQDLAAG---GKVTLTITNGS-----ERTVELTLKADGTLQPSDGNAY 987
Qy 831 KIGVKTTTL---TTTNANGKATNFSTTDNDALVNAKDAENLTLAKEIHTTKGTADTAL 887
```


Db 988 SYNAVGVIGWTEFFANGKSLVTVATQTDKAGNVSLPSSDGTATVLE-----APTADK-- 1042
QY 888 QTEKVKKDGATDDETTTVGK-----DGTQNGKTVNTLKLKGE-----NGLTVATNKO----- 934
Db 1043 -----QATGOEDAASIVPSLGSDDIDGTVSFVTSIOSLPTNGTLFLGSAVAVANQISITLA 1097
QY 935 --GTVTFGINTQSLGKAGDST-----TLNKDGLSIKKPNASNEQIQVGADGVKFAKVDKGN 988
Db 1098 DAGSLTFKPPNNW---GDTSFKYVAMNDGLADPTPAT-----LAIKVCSCND 1143
QY 989 STGI-----DGTSRITKDOIQTGANGSLDTTKPHLTCKDKLVKGEVETNTGINA 1038
Db 1144 APTLTISNDVAVSEGLPSGLKDEL---GEDGQKQADDEVVARGVITVGDVDSNLTVSL 1200
QY 1039 G---GKKITNIOQSDIT-----ONSDAVTGGRVYDLKTELESKINSAAK 1080
Db 1201 GWTGADVKSVSDGVKVEWSWDVSGSKTLIGYAGSKDNV-----LEVKLTPVGS-- 1252
QY 1081 TAQNSLHEFSVADQNHFTVSNPYSYDTSKTSVDITFAGENGITTKVKNKGVVRVIGIDQ 1140
Db 1253 -----KGDWGEVTLKQPLDHADTQ-----EDTIDLKINVSVDGALS 1292
QY 1141 TKGLTT-----PKLTVG-----NNGKGIVID-----SKD 1165
Db 1293 TKPLTTTVEDDAPLAGMPDAVSAKTDIPKVLTLGLFDLTQHSATANQRELDLDFKITAK 1352
QY 1166 GQNTITGLSNTLANVTNDGAGHALSOG-----LANDTD-----KTRAA 1203
Db 1353 GFKATSLSALIAADVASTAGLVAGSNPYNHLANEVDFRFDNGSKASEELIVKLDPG 1412
QY 1204 SIGDVLNAGNLNGEAVDFVSTYDTPFDIGNATTAKVYDTSK----- 1250
Db 1413 TVAYGANIKFSQMFGEKESGV-----VEFYRGGLVLISTLTFSDDAAGEYQNQFVQOG 1467
QY 1251 --TSKVYVDVNDKTEVTSO---KKLGKVTTLTKTSANGNATKFS---AADGDALVKA 1303
Db 1468 GFTMTWIKATNNGNTNISNSDFTVKSIEFLGTTTPOATAYSGTVAPOWGADGK----- 1523
QY 1304 SDIATHLNTLAGO---IQAKGASQASSAS-----YVDADGNKVIY-----DSTDK-K 1348
Db 1524 -----HIDLLGSNEVLMTATGEISISSKAGLPNTLYGEASDGLVFRLEFTPATGKWE 1378
QY 1349 YQVND-----KGQVD-KNKEVAKOKLVAQAOTPDGTLAQMNKVSINK--EQVNDANK 1399
Db 1579 FFQYQNMRLPGDGDIDFQKVVAD-----GDSSLGSFATKPLVTPVQSVSSGNA 1630
QY 1400 KQINEDNAPIKLENAKDTKNAAVTVGDLN-----AVAQ--TPLTFAGDTGTTA 1450
Db 1631 EGGNLSHTVT---LSDQTHEATQYFAIOGSGGANPASSDWGIAQFSNGVTYNSATGKIT 1688
QY 1451 KKLG-----ETLIK-GGQTDNKLTDNNIGVVAGTGDFTVKLAKD 1490
Db 1689 VPAGVSGFTVTLTVNDRLVETTELTVTVGGONGTGLINDRAPTA----- 1736
QY 1491 LTNLSNVNAGGTIDEKG-----ISFVDANQQA-----KANTPVLSANG-----LDLG 1533
Db 1737 -----AGGHSAGQEDKQVLTWAQFGASDDQAVTDLGVKINT--LPLNGTLEVINA 1786
QY 1534 GKRIISNIGRAVDNDVAVN-----FKQFNEVATVNN-----LNNQNSGASLPFVVTD- 1581
Db 1787 GQWVAVSAGALISYDEVNAGHLRFVPLGNLSESSNVNNGGALLGNLKGDYASIGYQISDG 1846
QY 1582 ---ANGK-----PINGTDGKPKAIGKADGKYHYHANAGVP----- 1614
Db 1847 ANLSNGSKLVIDVAADVAKTVNVT-----LTGNIPSYSPSPSPASITTELFQSGN 1898
QY 1615 VKDGPITDADKLANIAHGKPL--DAGHQVVASLGGNSDAI-----TLNLIKSTLP 1665
Db 1899 FNKTAFCITNS--LADSSSQEQVLGNSGNDYIVSTHGGGDLLYGOGNDVLVGGNAT-- 1954
QY 1666 QIDTPNTGNAN-----AGQAQSLPSLS-----AAQQNAASVKD-----V 1700

Db 1955 QGDSLYGGDGNVLYSGLSGSDGLYGGAGTDAIVLTGNRADYTIKKKSGSYSVNDRWFDFSV 2014
QY 1701 LNVGFNLQTNHNOVDFVKAYDTVNFVNGTGCADITSVRSADCTMSNITVNTALAAATDDGN 1760
Db 2015 TEKGAGITKALHDVEYVQFADGIYQLNQSGELTLVQP---TVVDYPLELSASLSDRDGS 2071
QY 1761 VLIKAKDGFYKADDLMPNGSLKAGKSASDAKPTTGLSLVNPAGKSGTGDVALNLSK 1820
Db 2072 EVL-----DSLVLSG-----MPKGAELYSGNQLLGTG-- 2099
QY 1821 AVFKSKDCTTTTSSDGLSIOQKD--NSSITLSKDGCLNVG-----GKVISNVG---K 1868
Db 2100 -----VDGKLTLLTAGSLWSASALDKVLSGLTLRVPSSAGQVLDKVEATAKEVGTQDTS 2154
QY 1869 GTKDTDAANVQQLNEVRNLLGLGNAGNDN-----ADGNQVNIADIK 1909
Db 2155 SASDEDSIRLSYFNATE-----GEPGNQNTFGSEHNIVVGDLDGSLVLPQNYNIAFMV 2209
QY 1910 KDPNS-GSSSNRTV-----IKAGTVLGGKGN---NDTEKLATGGVQVGVKDKG 1953
Db 2210 DSSGSGSDAMKTIKSOLAQOVFASLKSSAGTGDGAGKVNIFLVDFOVNKSISVDL-KDP 2268
QY 1954 NANGDLSNVWVKTKQDCKKALLAT-YNAAGQT--NYLTNNPAEADIRNEQGIKPFPHV 2010
Db 2269 DALNQLQNI-IKTM-DGSEKNGGATXYEDVKTANMFKSDFAKL--NTNAKNVAYF-IT 2323
QY 2011 DGNQEPVVOGRNGIDSSASCKSHSVAIGFOAKADGEAAVAIGR---OTQAGNOSIAIGDN 2066
Db 2324 DGQ-----PNTFNSESNPSIGYQWEGKWSAVYLSQIVTGYEQGEYVLDG-- 2370
QY 2067 AQATGQDSIAIGTGNV--VTGKHSGAIG-----DPSTVKADNSYSVGNNNQ----- 2110
Db 2371 -----RLVITTTTQVYGVDPGRWKGVNRPDKGKGYEAVTLSGDRD---GNSDQNARDGYA 2421
QY 2111 -FTDATQT-DVFGVGNNTVTE-----SNSVALGSNSAISAGTHAGTQAKKSDGTAGTT 2162
Db 2422 LLTAAGVTYDAIGIGDQLSVDDLLKGFSDGKVLGTGISADGLANAILGTPVSNLPGTDRLD 2481
QY 2163 TTAG---ATGTVRKGFAGQATAVGAVSVGASGASERRIONVAAAGEVSATS-TDA 2209
Db 2482 GGAGDDLLFGDAVHFAGINGEGYAAV-----KQYVAGKLSAGSVTDA 2523

Search completed: September 13, 2001, 12:57:20
Job time: 890 sec

us-09-361-619-11.rspt

Thu Sep 13 14:18:21 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 18125.9 Seconds
(without alignments)
3635.974 Million cell updates/sec

Title: us-09-361-619-5
Perfect score: 6972
Sequence: 1 ccattgatattggcagggtgt.....ccgtgacctgcaggtcgac 6972

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaton Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES             source
  Location/Qualifiers
    1..1101
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone_lib="RPCI-98"
      /clone="BACR08K10"
      /note="end : TET3"

BASE COUNT          201 a   64 c  131 g   202 t   503 others

ORIGIN
Query Match          0.8%; Score 56.8; DB 219; Length 1101;
Best Local Similarity 20.3%; Pred. No. 0.0012;
Matches 96; Conservative 174; Mismatches 202; Indels 0; Gaps 0;

QY  102 acatgattaccattgttaactgttgccatcattaccataatgtagtaacgcat 161
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  396 ATATAWAANWNTTTTTTANAAWAAATAATTTWAAWAAWAAATTTWAAWAA 455
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  162 gtaacgcattgtgaaatcatctgcccctttgtgtatcatgaatgaatattat 221
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  456 ANAWTAWNTTAAWAAAAAATAATTTTWTTTTAAWTAATWATWNTTAA 515
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  222 gattgatctgattgtatcagaatggtgatgtatgatgatgcctacgattgat 281
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  516 AAAAAAATAAATAAATAAATAATTTWNTTTTAAWATAATTAANCAAAWYHY 575
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  282 ttgggttaactctatgtatgatattttgaaactaatctattgacttaacacc 341
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  576 TTYVHYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 635
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  342 atatggttaattagcataatgtaggcttttggtaaaatcacatgcgaattgt 401
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  636 TAAAYYYYTCMYHYHYHHHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 695
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  402 ctactgttactaccatgctgaatgacgacatcccaatcacagatcattcaatg 461
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  696 HTCHHCYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 755
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  462 ttgtatagcaccatttaccctaatatttcaatcaaatgctatgtcagcatgat 521
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  756 TWYAWAHAMWVHHHAHYAAAAAATAATTHVYVYVYVYVYVYVYVYVYVY 815
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  522 ttttttaaggtaaacaccatgaatcacatcatataaagtcattttaaaca 573
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  816 WHYYTAYTCWTTHWNTWTHWYHHHTTHWHTTHWAWWHTTWCWWMWHA 867
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :

```

```

RESULT 6
LOCUS   CNS02EOD/c 681 bp DNA GSS 13-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 262H14 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL193990.1 GI:7832096
VERSION   AL193990
KEYWORDS  GSS: genome survey sequence.
SOURCE    Tetraodon nigroviridis.
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

```

REFERENCE
AUTHORS  Acanthomorpha: Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 681)
Roeest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 681)
Roeest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 681)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

```

```

FEATURES             source
  Location/Qualifiers
    1..681
      /organism="Tetraodon nigroviridis"
      /db_xref="taxon:99883"
      /clone="262H14"
      /clone_lib="G"
      /note="Genoscope sequence ID : C0AG262DD07LPI-end : T7"

BASE COUNT          74 a   105 c  170 g   282 t   50 others

ORIGIN
Query Match          0.8%; Score 56.4; DB 220; Length 681;
Best Local Similarity 41.1%; Pred. No. 0.0013;
Matches 172; Conservative 26; Mismatches 221; Indels 0; Gaps 0;

QY  3131 accatgatcgcgttaacaaaccagtaagtggtatgtatgtgaatgtggtatgata 3190
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  450 ACTTWTTCASAGACACGGAGACRATTTTWTWTATWATTTTGTCDACRACAACWACAAC 391
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  3191 accattcatctaaaggcactgatgacaataaaacttggcgtcaaaaccaccacaaactg 3250
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  390 AAAAAAACRACWACTACAAACAAWACAACACGGCGGCAACAAACACGACGACGACGAC 331
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  3251 aacaaaaaagtctaatgttaacagcaacttaactttaattgttaactctgtgtgaa 3310
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  330 GACACAAAAACGACRACGCGGACGACGACGACGACGACGACGACGACGACGACGAC 271
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  3311 gatgccttgttaacgcaagacatgcgcgaaatctaaacacccctagccaaaggaatt 3370
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  270 GACGCAACAAAAACRCAACRACGCGGACGACGACGACGACGACGACGACGACGACGAC 211
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  3371 cacacacaaaggcacagcacacgcctcctcaaacctttaccgttttaaaaggtagat 3430
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  210 AACACRCAACAAACNACGACGACGACGACGACGACGACGACGACGACGACGACGAC 151
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  3431 gaaaataataatgtgtgacgcacacgcctacacgcgttggtcaaaagacgcgaataat 3490
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  150 GACACAAACACRCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACRAC 91
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
QY  3491 caagtcaacacccctaacctcaaaagtgtaaacgcttcttaattattaaaccacacaaaa 3549
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :
DB  90 RGCRCGCAACAAACAAACAAACRCAACGCGGACGACGACGACGACGACGACGACGACGAC 32
    ||| : : : : ||| ||| : : : ||| ||| : : : ||| : : : ||| :

```

```

RESULT 7
CNS00EVL/c 1101 bp DNA GSS 04-JUN-1999
LOCUS     Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit

```


[illegible]

RESULT	9	
CNS00DKY/c		
LOCUS		
DEFINITION	CNS00DKY 928 bp DNA GSS 04-JUN-1999	
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #	
DEFINITION	BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit	
DEFINITION	fly), genomic survey sequence.	
ACCESSION	AL071865	
VERSION	AL071865.1	
KEYWORDS	GI:4948170	
SOURCE	GSS.	
ORGANISM	fruit fly.	
ORGANISM	Drosophila melanogaster	
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
ORGANISM	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
ORGANISM	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 928)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-JUN-1999), Genoscope - Centre National de Sequencage :	
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	
JOURNAL	- Web : www.genoscope.cns.fr)	

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/Drosophila_Bac.htm.

FEATURES	source
round at	recept//prophila/locos/prophila_bac.htm.
Location/Qualifiers	
1. .926	
/organism="Drosophila melanogaster"	
/db_xref="taxon:7227"	
/clone_lib="RPC1-98"	
/clone="BACR37A24"	
/note="end : T7"	
BASE COUNT	262 a 70 c 84 g 321 t 191 others
ORIGIN	

```

Query Match      0.88; Score 54; DB 219; Length 928;
Best Local Similarity 31.2%; Pred. No. 0.0062;
Matches 96; Conservative 72; Mismatches 140; Indels 0; Gaps 0;

Qy 103 catgatttaacattgttttaatactgttgccatcatcattatgtaacgcattag 162
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

[illegible]

RESULT	10
CNSOOFUH/c	
LOCUS	CNSOOFUH 996 bp DNA GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR1021 of RPECI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL071063
VERSION	AL071063.1 GI:4951105
KEYWORDS	GSS.
SOURCE	fruit fly, Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 996)
REFERENCE	Genoscope.
AUTHORS	
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial *EcoRI* digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	source	Location/Qualifiers
BASE COUNT	383 a	164 c
ORIGIN	171 t	197 others
		1. .996
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_lib="RPCI-98"
		/clone="BACR31021"
		/note="end : TET3"

Query Match	0.8%	Score 53.4	DB 219	Length 996
Best Local Similarity	36.0%	Pred. No. 0.0093		
Matches 125	Conservative	52	Mismatches 170	Indels 0
				Gaps 0

Matches	167;	Conservative	53;	Mismatches	241;	Indels	0;	Gaps	0;
QY	83	tgatgattgaacacgtgcacatgattgaacactgttgaatactgttgcccatcattacca	142						
Db	1100	TTATTGTTTTTAAATWAAAATAAATTAATWTWTATATTWAAATTTTAAATTTTWWTATW	1041						
QY	143	taatttagtaacgcattagtaacgcatttgtaaaaatcattcgcccctttatgtgat	202						
Db	1040	TTWATWATTAAATATWATTTTTTAAAAAATAWAAAWTTTWTTTTAATTWAAATTTTAAATW	981						
QY	203	catatgaagaatattatgattgtatctgatttatgtatcagaagtggatgctatatg	262						
Db	980	TTTTTAAATTTTTTWWATTTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	921						
QY	263	atgatgccacgagtgagggtggttaacctcacctcattgatattgatatttgaaacctaa	322						
Db	920	AATATAAAWAATAATTTTAAATATTWTTTTAAATAAAATTTTWTTTTATTTTATTTAT	861						
QY	323	tctattgacttaaatcaccatattggtttataattagcataatggtagcctttttgtaaaa	382						
Db	860	TTWATTTAAATTTTAAWATWATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	801						
QY	383	atcacatcgcaatattgttcctactgttactaccatgcttgaatgacgatccccaatcca	442						
Db	800	WATTTTWTTTTTTTTAAWTTTAAATTTATATATAATAATAWAAWAAAAAATAAAAAA	741						
QY	443	gattcattcaagtgatggtgttgatacgcacattccaccattatttcaatcaaatacg	502						
Db	740	ATAAAAAATAAATTTTTTTTTTTTTTTTATATAASVMTWKVMASSCAATVTWANAATA	681						
QY	503	cctatgctgcagcatgattcatttttttaaaggtaaacaccacct	543						
Db	680	AMAATSVCASBSVWTTATATTTTATTTTASAASAAWAKAWAAT	640						

RESULT 12
AQ946120/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
COMMENT

FEATURES
source

```

AQ946120      641 bp      DNA              GSS          27-JAN-2000
Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-46J23, DNA sequence.
AQ946120
AQ946120.1   GI:6769385
GSS.
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 641)
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other_GSSs: Sheared DNA-46J23.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.
Location/Qualifiers
locus_tag . . . 641
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
```

Query Match 0.8%; Score 53; DB 219; Length 1101;
Best Local Similarity 36.2%; Pred. No. 0.012;

ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 958)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES	Location/Qualifiers
source	1..958
	/organism="Drosophila melanogaster"
	/plasmid="pBelobAC11"
	/db_xref="taxon:7227"
	/clone_lib="DrosBAC"
	/clone="BACN17A05"
	/note="end : SP6"

BASE COUNT	94 a 116 c 93 g 495 t 160 others
ORIGIN	

```

Query Match      0.7%; Score 52.2; DB 219; Length 958;
Best Local Similarity 39.3%; Pred. No. 0.019;
Matches 171; Conservative 33; Mismatches 231; Indels 0; Gaps 0;

QY 3115 caccacgcacagtcacccatgatacgcgttaacaacaacagtaaagtggatatgatgt 3174
    || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 567 CAAAAGGCGCAAGAGAGAGAACCCAAACCAACAMAAGAAGCRAARAGAGAACAAAA 508
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
QY 3175 gaatgtagatcaaacattcatctaaaggcgactgatgacataaaaaacttggcgt 3234
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
Db 507 AAAAGCGMAARMWAACCAAMWACCACGCCACCAAACACACAACAAAAA 448
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
QY 3235 caaaaccaccaactgaacaaaagaatgcttaattggttaatacagcaactaactttaatgt 3294
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 447 AAAAAAAAAAAAAAAAAAACAACGAGAAAGCAACACAAAGCAAGAAAGMAAAMGA 388
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
QY 3295 taactctagtgatgaagatgcccttgtttaacgcgcaaagacatccgcgaaatctaaacac 3354
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
Db 387 AAAAAAAAAAGAAAAACCAANMAAAACGARAAANMARAANAAMAAAMAAAGCAACAAAA 328
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
QY 3355 cctagccaaggaaattcacaccaccaaaggcacagcagacacccctcacaaaccttac 3414
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 327 GAACCAAAAAAAAAAARCCAAAAACAAAMAAACAAAMAAAGACRAMAAAMAAACGAAAC 268
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
QY 3415 cgttaaaaaggtagataaaataatgctgtatgcgcgcaacgccatcacctggggtca 3474
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 267 MAIAAAAAAAAAAGCAAGAAAAAAMAAAAAAAAAAAAAAAAAAGAGAAAAACAARRAM 208
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
QY 3475 aaegaacgcaaatataatcaagtcaacacctcaactcaaaagtgtaaaacggtcttaatat 3534
    || | : : : : : ||||| : : : : : ||||| ||||| ||||| ||||| |||||
Db 207 AGRAAAAMRAAARAAAMAAAAAAMAAAAAACMAAAAAAACCAAGCAACAAAAAGAA 148
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
QY 3535 taaaccgcacaaaaa 3549
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 147 AAAAAAAAAAAAAA 133
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

```

RESULT 15
BF278771/c 780 bp mRNA EST 07-MAR-2001
LOCUS BF278771
DEFINITION GA_Eb0033M02f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cdNA clone GA_Eb0033M02f, mRNA sequence.
ACCESSION BF278771

```

VERSION      BF278771.1  GI:11209757
KEYWORDS     EST.
SOURCE       Gossypium arboreum.
ORGANISM     Gossypium arboreum
Eukaryota;  Viridiplantae;  Streptophyta;  Embryophyta;  Tracheophyta;
Spermatophyta;  Magnoliophyta;  eudicotyledons;  core eudicots;
Rosidae;  eurosids II;  Malvales;  Malvaceae;  Gossypium.
1 (bases 1 to 780)
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCATATAGGG
High quality sequence stop: 200.

FEATURES             source
1..780
    /organism="Gossypium arboreum"
    /strain="AKA"
    /cultivar="8400"
    /db_xref="taxon:29729"
    /clone="GA_EB0033M02f"
    /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
    /tissue_type="Fibers isolated from bolls harvested 7-10
    dpa"
    /lab_host="E. coli"
    /note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      137 a      63 c      137 g      443 t

Query Match      0.7%;  Score 52;  DB 146;  Length 780;
Best Local Similarity 44.0%;  Pred: No. 0.02;
Matches 220;  Conservative 0;  Mismatches 280;  Indels 0;  Gaps 0

Qy 2990 aaagacaaatccaacgctgccagcattatgatattattaaataacaggctttaacctaaaa 3049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 696 ACAACCAACACAAACTAGATAATAAAAAAATAAAAAATAAATACAAAAAATAAAAA 637
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3050 aataatacaaccccatgactttgtctccacttatgacattgttgactttgccaatgac 3109
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 636 ATCAAAATATCGAAAAAATAATTTATTAACAGCAAAACAAATGAATAACATTACACTAA 577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3110 aatgccaccgcgcacagtagcaccatgataccgcttaacaaacagtaaggtggtatat 3169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 576 AAAACACCCAGAACAGTATACAAAAACACCCAAATACAAATATAATATAATAATAA 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3170 gatgtgaatgtgatgatacaaccattctcaacaggcactgatgacataaaaaactt 3229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 516 AAACTAACACTCAATTAACACACACACAATAATACAACGTATACATAACAAAAACAATAACAAA 457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3230 ggcgtcaaaaccccaactgaacaaaaacagtgtctaattggtaatcacgcaactaacttt 3289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 456 AAAGCAAAACAAAAGAAAACAAATAATAACACAAATCAAAATTCGTATAAAACACAAAAA 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3290 aatgttaactctagtgaagatgcccttgttaacgccaagaagacatcgccgaaaatcta 3349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 AAAAGAAACAGAAATAAACAAAAACATTTTCATCTTTCATCCAAATACAAATAAAAAAAA 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3350 aacacccctagccaaggaattcacaccacccaaggccacagcagacacgcctcacaaacc 3409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 AAAACCAAAATACAAAAAAGAAAAAATAAAAAAATAAAAAAATAAAAAATCAAAAAAAAAT 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3410 ttaccggttaaaaggttagatgaaataataatgtctgatgcgcgcacgcgcacccgtg 3469
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 CACACCTAAAAAATTTAAAAATATATAAAAAAAGAAAGTAACATTAAACAATCCAAAAA 217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT	15		
LOCUS	BF278771/c		
DEFINITION	BF278771	780 bp	mRNA EST
ACCESSION	BF278771	GA_EB0033M02f	Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA_EB0033M02f, mRNA sequence.

Qy 3470 ggtcaaaagaacgcaataa 3489
||| | || |||
Db 216 ACAAAAATCAGTCACATTA 197

Search completed: September 13, 2001, 03:33:34
Job time: 53636 sec

Db 1 CCATGATATGGCAGGTGTGCTGCCGTATGATGGCGATGACACCCCAATTTGCC 60
Qy 61 catabctgtacgatttgacatgtgatgatatttaaacatgtgacatgtatgaattgaattgttt 120
Db 61 CATATCTGTACGATTTGACATGTGTATGATTTAAACATGTGACATGATTTAACTGTTT 120
Qy 121 aatactgttgcctcattaccataaatttagtaacgcatttagtaacgcattttgttaaaat 180
Db 121 AATACGTGTCACATACCATTAATTTAGTACGCATTTAGTAAAGCATTTGTAAANAT 180
Qy 181 cattgcccctttatgtgtatcatatgaatagaatattatgtatgtatctgtatttgc 240
Db 181 CAITTGCCCCCTTTATCTGTATCATATGAATAGATATTTATGATTTGATCTGATTTATCT 240
Qy 241 atcagaatggtgatgtatgatgatgctacagattgatttgggttaactcaactctatg 300
Db 241 ATCAGAAATGATGATGATATGATGATGCTACGAGTTGATTTGGGTTAATCACTCTATG 300
Qy 301 atttgatatatttgaactaaactctattgacttaactcaacatatgtgtataatttagca 360
Db 301 AATTGATATATTGAACTTAATCTATTGACTTAATCAACCATATGTTATATTTAGCA 360
Qy 361 taatggtagctttttgtaaaaatcaacatgcgaatatattgtctactgttactaccatgct 420
Db 361 TAATGGTAGGCTTTTGTAAAAATCACATCGCAATATTGTTCTACTCTTTACTACCATGCT 420
Qy 421 tgaatgacatcccaatcaccagattcattcaagtgtgtgtttgtatagcaccattta 480
Db 421 TGAATGACGATCCCAATCACAGATTCATTCAAGTGTGTTGTGTATAGCGACCATTTA 480
Qy 481 ccctaattatttcaatcaaatgctatgtcagcatgtatcatcttttttaaggttaaacccac 540
Db 481 CCCTAATTATTCAATCAATGCCATGTCTCAGCATGTATCATTTTAAAGGTAACCCAC 540
Qy 541 catgaatcaatctataaagtcatctttaacaaagccacagcacatttatgcagtgcc 600
Db 541 CATGAATCACATCTATAAAGTCATCTTTAAACAAAGCCACAGGCACATTTATGGCAGTGGC 600
Qy 601 agagtacgcaaatccacacagcac- ggggggggtagctgtctacagggcaagtgggca 659
Db 601 AGAGTACGCCAAATCCACAGCACAGCGGGGGGTAGCTGTCTACAGGGCAAGTTGGCA 660
Qy 660 gtgtatgcaactctgtagctttgcccgtattgcccgcgtcgtgtcctgtatgctggtgca 719
Db 661 GTGTATGACATCTGTAGCTTTTCCCGTATTGCCGCTCGCTGTCTCCTGTATCGGTGCA 720
Qy 720 cgtcagtggtgagtgcttatgtctaaaaaagataccaaacatatcgcattgtgtgaac 779
Db 721 CGCTCAGTGGCAGTCTTATCTCAAAAAAAGATACCAACATATCGCAATTTGGTGAAC 780
Qy 780 aaaaccagccaagacgtctcagggcactgccaggcggcgtgatacggaccattgctattg 839
Db 781 AAAACAGCCCAAGCGCTCAGGCATGCAAGCGGAGCGGTGATCGAGCCATTTGCTATTG 840
Qy 840 gtgaaaaatgctaagcagcaggggtcaagccatcgccatgtagtagtaataaaactg 899
Db 841 GTGAAATGCTAAACGACAGGGCGGTGAAGCCATCGCCATCGGTAGTAGTAAATAAACTG 900
Qy 900 Lcaatgggaagcagtttggataagaatgtagtaccgatgctacgggtcaagagtcacatcgcca 959
Db 901 TCAATGGAAGCAGTTTGGATTAAGATAGTACCGATGCTACGGGTCAAGAGTCCATCGCCA 960
Qy 960 tcgggtggatgtgtaaggtcgtgtagtgcctcgattgccaatcggttagtgatgaacttac 1019
Db 961 TCGGTGGTGTATTAAGGCTAGTGTGTATGCTCGATTGGCATCGGTAGTATGATGACTTAC 1020
Qy 1020 atttctgtatgacgatggttaactcaacatccgaagggttactctgatttaacgatctta 1079
Db 1021 ATTTCCTTGATCAGCATGCTGTATCTTAACATCCGAAGGTACTCTGATTAAACGATCTTA 1080
Qy 1080 ttaacggccatgcagttatataaagaaatcacgaagctcaaaagataatgatgtaaaatata 1139
Db 1081 TTAACGGCCATCGCATTTAAAAAGAAATACGAAGCTCAAGGATAAATGATGATAAATATA 1140

Qy 1140 gacgcacaacgcgaagcggacacgcagctactgtcagttgggagccatgtcatatgcacagg 1199
Db 1141 GACGCACAACCGCAAGCGGACACGCGCAGTACTGTAGTGGAGGCCATGTCTATATGCACAGG 1200
Qy 1200 gtcaatttttccaacgccttttggtagacggtggaacagctaaaggtgcctattccttggcag 1259
Db 1201 GTCAATTTTCCAAACGCTTTGGTACACGGGCAACAGCTTAAAGTGGCTTATTCCTTTGGCAG 1260
Qy 1260 tgggtctgcccacacagccaggggccaatctacaaatcgctattggttctgtagcaaat 1319
Db 1261 TGGGCTTTGCCGCCACACGCCGAGGCCAATCTTACAAATCGCTATTGGTTCTGATGCAACAT 1320
Qy 1320 ctatgctgttgggagcagatagcccttgggtgcagggtactgtctcagctacagggcagta 1379
Db 1321 CTAGCTCGTTGGAGCGGATAGCCCTTGGTGCAGGTACTGCTGCTCAGGTACAGGGCAGTA 1380
Qy 1380 ttgcccctagggtcaaaagtctgtgtcactcaagtgataataattctagaccgacctata 1439
Db 1381 TTGCCCTTAGGTCAAGGTTCCTTTGTCTACCTCAGAGTGAATAATTAATCTAGCCGGCTATA 1440
Qy 1440 caccaaaataccagcagcactagaccccaagtttcaagccaccaataatacgaaggcggttc 1499
Db 1441 CACCAAAATACCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGCGGCTC 1500
Qy 1500 caatttccattgtgtagtaactctatacaaacgttaaaatcatcaatgtcgggtcaggtgtta 1559
Db 1501 CACTTTCCATTGTGTAGTAATCTCTATCAAAACGTTAAATATCAATGTCTGCTGCAGGTGTTA 1560
Qy 1560 ataaaacgcagtcggtcgaatgtggcacagctagaagcgtggtgaagtggtggtcgaaggagc 1619
Db 1561 ATAAAACCGATCGGTCAATGTGGCACACCTTGAAGCGGTGTAAGTGGCTTGAAGGAC 1620
Qy 1620 gtagaattacttttcagggtgatataacagttactgacgtaaaaaataggttttggataata 1679
Db 1621 GTAGAATTACTTTTTCAGGGTGTATGATTAACAGTACTGACGTAAAAATAGGTTGGATAATA 1680
Qy 1680 ctttaactattaaagtggtgcagagacaaagcatttaaccgataataataatcgggtgtg 1739
Db 1681 CTTTAACTATTAAAGTGGTGCAGACCAACGATTAACCGATAATATATTCGGTGTGG 1740
Qy 1740 taaaagggctgataatagttggtctgaaagttaaactgtctaaaaactttaaacaaactta 1799
Db 1741 TAAAGAGCGCTGATATAGTGGTCTCAAAAGTTAACTTGTCTAAAACTTTAAACAACTTTA 1800
Qy 1800 ctgaggtgaatacaactacataataatgaatgcacacacacagtttaaggttagtagtagta 1859
Db 1801 CTGAGGTGAATACAACTTACATTTAAATGCCACACACAGTTAAAGGTAGGTAGTAGTA 1860
Qy 1860 gtactacagctgaattattgagtgatgtttaacctttaccagcccaatacagggcagtc 1919
Db 1861 GTACTACAGCTGAATTTATTGAGTGTAGTTTAACTTTACCCAGCCCAATACAGGCAGTC 1920
Qy 1920 aaagcacaagcaaaacgcgtctaaggcgttaagtgggtgaagtttactaaataatgcgaaa 1979
Db 1921 AAAGCACAAGCAAAACCGTCTATGGCGTTAATGGGTGAAGTTTACTTAATAATGCGAAA 1980
Qy 1980 caacagcagcaaatcggcactactcgtattaccagagataaaattgcttctgctcgagatg 2039
Db 1981 CAACAGCAGCAATCGGCACCTACTCGTATTTACCAGAGATAAAATTTGGCTTTGCTCGAGATC 2040
Qy 2040 gtgatgttgatgaaaaaacaagcaccataatttggatataaaaaaacacttaaaagtgggtagtg 2099
Db 2041 GTGATGTGTGATAAAACAAAGCACCATATTTGGATAAAAAACAACCTTAAAGTGGGTAGTG 2100
Qy 2100 ttgcaattacatagacaaatggcatttgatgcaggttaataaaagatcagtaactcttgcga 2159
Db 2101 TTGCAATTTACCATAGACAAATGGCATTTGATCAGGTAAATAAAAAAGATCAGTAACTTTGCCA 2160
Qy 2160 aaggtacagatgcttaacgatgcggttaccatcgaacagctcaaaagccgcaagcctactt 2219
Db 2161 AAGGTACAGTGTAAACGATGCGGTTTACCATCGNACAGCTCAAAAGCCGCCCAAGCCTACTT 2220

Db 4381 CGGTTGACTTTGTCTCCACTTATGACACCGTCAACTTTTGGCGATGGCAATGCCACACCG 4440
Qy 4440 ctaaggtgacctatgatgacacaaagcaaaacagtaagaatggtcttatgctgaatgtg 4499
Db 4441 CTAAGGTGACCTATGATGACACAAAGCAAAACCAAGTAAGTGGTCTATGATGTCATGTGG 4500
Qy 4500 atgatacaaccattgaattgaagtaaaagataaaaccttggcgtataaaacccacacattgacca 4559
Db 4501 ATGATACAACCATTTGAAGTTAAAGATTAATAAACTTTGGCGTAAAAACCCACCATTTGACCA 4560
Qy 4560 gtactggcaaggtgtaataataattgacctaaagcaatacaagctactggcgatggcgtctg 4619
Db 4561 GTACTGGCACAGGTGCTAATAAATTTGCCCTTAAGCAATCAAGTACTGGCGATGGCGCTTG 4620
Qy 4620 tcaagggcagtgatcgtgtgctcatctataaacacotttatctggcgacatccaaaactggca 4679
Db 4621 TCAAGGGCAGTGATGCTGTGCTCACTTAACACACCTTATCTGGCGACATCCAAACTGGCCA 4680
Qy 4680 aaggggcaagcgaagcgaacactcagcaggttatgtggtgctgctgagcaataaaggtca 4739
Db 4681 AAGGGCAAGCCCAAGCAACTCAGCAGGCTATGCTGGATGCTGATGCAATGAAGTCA 4740
Qy 4740 tctatgacagctacgatacaagtaactatacaagccaaaatgatggcagcttgataaaaa 4799
Db 4741 TCTATGACAGTACCGATAACAGTACTATATCAAGCCAAAATGATGCGACAGTTGATAAAA 4800
Qy 4800 ccaaaaggtgcaaaagcaaaactggtgcgcaagcccaaacccacagatggcacattgg 4859
Db 4801 CCAAGAAAGTTGCCAAGACAACAACTGGTCCGCCCAAGCCCAACCCAGATGGCACATTGG 4860
Qy 4860 ctcaaatgaatgtcaaatcagtcatttaacaaagaacaagtaaatatgctgccaataaaagc 4919
Db 4861 CTCAAATGAATGTCAATCAGTCACTTAACAAAGAACAAGTAATGATGCCAATAAAAAGC 4920
Qy 4920 aagggcatcaatgaagcaacgccttctgttaaggacttaaaagccgctcttgataaca 4979
Db 4921 AAGGCATCAATGAAGCAACGCCTCTGTAAAGGACTTGAAGAAGCGCTCTTGATAAACA 4980
Qy 4980 aacccaaaacgcgcagtagtaactgtgggtgatttaaatgcgttgcacaaacacccgtga 5039
Db 4981 AAACCAAAAACGCCGACAGTAACGTGGGTGATTTAAATGCCGTGGCCCAAAACCCGCTGA 5040
Qy 5040 cctttgaggggatacagggcacaaaggctaaaaactggcgagactttgaccatcaag 5099
Db 5041 CTTTTCAGGGGGATACAGGGCACACCGCTAAAAAATCGCGGAGACTTTGACCATCAAG 5100
Qy 5100 gtgggcaaacagacaccaataaagctaaacgataataacatcggtgtggtgtagcaggtactg 5159
Db 5101 GTGGGCAACAGACNCCANTAAAGCTAACCGTAAATACATCGGTGTGGTAGCAGGTACTG 5160
Qy 5160 atggcttcaactgtcaaaactgcaaaagacctaaccatcttaacagcgttaatgcaggtg 5219
Db 5161 ATGGCTTCACTGTCAAACTTTGCCAAGACCTTAACCAATCTTAACAGCGTTAAATGCAGGTG 5220
Qy 5220 gacccaaatgtatgacaaaggcgtctctttgtaagctcaagcgtcaagccaaagcaa 5279
Db 5221 GCACCAAAATTTGATGACAAAGGCGGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAA 5280
Qy 5280 acacccctgtctgaatgccaatgggtgacccgtgggtgcaaggtcaatcaagtaatgtg 5339
Db 5281 ACACCCCTGTCTAAGTGCCAAATGGCTGGACCTGGGTGCGCAAGGTCAATCAGTAATGTGG 5340
Qy 5340 gcaagggcaaaaaagataccgagcgtcccaatgtacaacaggttaaacgaagtagcgaact 5399
Db 5341 GCAAGGCGACAAAAGATACCCACGCTGCCAATGTPACAACTAGTTAAACGAAGTACGCAACT 5400
Qy 5400 tgttgggtctgtgtaatgctgtgtaataacgctgacgcaatcaggttaaacattgccc 5459
Db 5401 TGTGGGTCTTGGTAATGCTGGTAAATGATTAACGCTGACGGCAATCAGGTAACAATTGCCG 5460
Qy 5460 acataaaaaagaccccaattcaggttcatcatcctaaccgcactgtcatcaaaagagcca 5519
Db 5461 ACATCAAAAAGACCCAAATTCAGGTTTCATCATCTAACCCGCACTGTCTATCAAAAGCAGGCA 5520

Qy 5520 cggttacttggcgttaaaagtaataacgataccgaaaaaacttgcactggtggtatatacaag 5579
Db 5521 CGGTACTTGGCGGTAAAGGTAAATACGATACCGAAAAACTTTGCCACTGGTGTATACAAAG 5580
Qy 5580 tgggcgtggaataaaagacggcaacgcgtacagcgttaagcaaatgttgggtcgaacaccc 5639
Db 5581 TGGCGGTGATTAAGACGCGCAACGCTAACGGCGATTTAAGCAATGTTGGGTCAAAACCC 5640
Qy 5640 aaaaagatggcagcaaaaaagccctgctcgccactataacgcgcaggtcagaccaact 5699
Db 5641 AAAAAGATGGCAGCAAAAANGCCCTGCTGCCACTTATAACGCCGAGGTCAAGCAACT 5700
Qy 5700 atttgaccaacaacccccgcagagcatttgacagaaataaatagaacaaggtatccgctctct 5759
Db 5701 ATTTGACCACAACCCCGCAGAGGCATTTGACAGAATAAATGAACAAGTATCCGCTTCT 5760
Qy 5760 tccatgtcaacatagggcaatcaagagcctgttggtaacaaggcgtaacggcattgactcaa 5819
Db 5761 TCCATGTCAACATGGCAATCAAGAGCCTGTGTACAAAGGGCGTAACGGCATTTGACTCAA 5820
Qy 5820 gtcctcagggcaagcaactcagtggcgataaggtttccagggcgaagcagatggtgaagcgc 5879
Db 5821 GTGCCTCAGCAGCAACTCAGTGGCGATAGGTTTCCAGGCCAAGGCAGATGTTGAAGCCG 5880
Qy 5880 ccgttgccatagggcagacaaaccccaagcaggcaaccaa tccatcgccatcggtgataaacg 5939
Db 5881 CCGTTGCCATAGCAGACAAACCCAGCAGGCAACCAATCCATCGCCATCGGTGATAAGC 5940
Qy 5940 caaagcagcgggatcaatccatcgccatcggtacagggcaatgtggtagcaggttaagc 5999
Db 5941 CACAAGCAGCGCGCATCANTCCATCGCCATCGGTACAGGCATGTTAGGCTGATAACAGT 6000
Qy 6000 actctgtgccaatcggcgaccccaagcactgttaagctgataacagttacagttcagttgggta 6059
Db 6001 ACTCTGTCATCGCAGCAGCCCAAGCAGCTGTTAAGGCTGATAACAGTTACAGTGTGGTA 6060
Qy 6060 ataacaacagtttacgatacgccactcaaaacgagatctcttgggtgggcaataaacaatca 6119
Db 6061 ATAACAACAGTAACTACCGATGCGCACTCAAAACCGATGCTTTGGTGGGCAATAACATCA 6120
Qy 6120 ccgtgaccgaagtaactcggttgccttaggttcaaaactctgccaactcagtcagtcagggcacac 6179
Db 6121 CCGTGACCGAAGTAAGTACCTCGGTGCTTAGTTCAAACTCTGCCATCAGTCAGCGCACAC 6180
Qy 6180 agcagggcacaaagcgaataatctgcggaagcagcaggtgaggtcagtcagtcagccacga 6239
Db 6181 ACGCAGGCACAAAGCCAAAATCTGACGGCACAGCAGGTACACACACACAGCAGGTG 6240
Qy 6240 caacgggtacggttaagagcttctggcaaaacgcggttgggtcggtctccgtgggtg 6299
Db 6241 CAACCGGTACGGTTAAAGGCTTTGCTGGACAAACGGCGGTGGTGGGCTCCGTGGGTG 6300
Qy 6300 cctcaggtgctgaacccggtatccaaaaatgtggcagcaggtgaggtcagtcagtcagccacga 6359
Db 6301 CCTCAGGTGCTGAACCGCTATCCAAAATGTGGCAGCAGGTGAGGTGAGTCAGCCACGCA 6360
Qy 6360 ccgagtcggtcaatggttagccagttgtcaaaagccaccccaaaagcattgccaaagcaacca 6419
Db 6361 CCGATCGGTTCAATGTGTAGCCAGTTGTACAAAGCCACCCAAAGCATTTGCCAACGCAACCA 6420
Qy 6420 atgagcttgaccatcgltatcccaaaaacgaaaaaagggccaatgcagggatttcatcag 6479
Db 6421 ATGAGCTTGACCATGCTATCCCAACCAAAACGAAATTAAGGCCAATGACGGGATTTTCATCAG 6480
Qy 6480 cgaatggcagtggttccatggcacaagcctacattcctggcagatccatggttaccgggg 6539
Db 6481 CGATGGCATGGGTCCATGCCACAAGCCCTACATTCCTGGCAGATCCATGTTGTTACCGGGG 6540
Qy 6540 gtattgccaccccaacacaggttcaaggtgcggtggcagtgaggactgtcgaagctctcgata 6599
Db 6541 GTATTGCCACCCCAACCGGTCAAGGTGCGGTGGCAGTGGGACTGTCTGAAGCTGTGCGATA 6600

Qy	4362	tgaagcgcaatggtgaagcggttgaccttggctccaacttatgacacccgtcaactttgcgcg	4421
Db	4266	TGCAAGCAATGCTGAAGCGGTTGACTTTGCTCTCCACTTATGACACTGCTCAACTTTGC	4325
Qy	4422	atggcaatgcacacccacgcctaaagtgacactgatgacacaaagcaaaacacagtaaatgg	4481
Db	4326	ATGGCAATGCCACCGCTTAAGTGGACCTATGATGACAAAGCAAAACCAAGTAAAGTGG	4395
Qy	4482	tctatgatgtcaatgttgatgtatatacaacacctgaagtttaaagataaaaaacttgcgttaa	4541
Db	4386	TCTATGATGTCAATGTGGATGATACAACCATTTGAAGTTAAAGATAAAAACTTGGCGTAA	4445
Qy	4542	aaacccacacatigacagtaactgagcagggtgctaataaatttgccttaagcaaatcaag	4601
Db	4446	AAACCCACCATTTGACCACTTGGCAGTACTGGCAGAGTGTCTAAATAATTTGCCCTTAAGCAATCAAG	4505
Qy	4602	ctactggcgatgcgctgttcaagggccagtgatatgttgctcatctaaacaccttatctg	4661
Db	4506	CTACTGGCGATGCGCTTGTCAAGGCCAGTGATATGTTGCTCATCTAAACACCTTATCTG	4555
Qy	4662	gcgacatccaatactgccaaaggggcgaagcccaagcgaaacaaactcagcaggctatgtggatg	4721
Db	4566	GCGACATCCAACTGCCAAAGGGGGAAGCCAAAGGAACAGCTCAGCAGGCTATGTGGATG	4625
Qy	4722	ctgatggcaataaggtoactctatgacagtaacgataacaaagtactatcaagcccaaaaatg	4781
Db	4626	CTGATGGCAATAAGGTTTCATCTATGACAGTACCGATAAACAAGTACCTATCAAGCCAAAAATG	4685
Qy	4782	atggcacagttgataaaacaaagaagttgccaagacaaactgctgcccaagcccaaaa	4841
Db	4686	ATGGCACAGTTGATAAACCAAGAAGTTGCCAAAGACAAACTGTGTGCCCAAGCCCAAA	4745
Qy	4842	cccagatggcacattggctccaatgaaagtccaaatcagctcatatacaagaacaaagtaa	4901
Db	4746	CCCCAGATGGCACATTGGCTCAATGAATGTCAAACTAGTCATTAAACAAGACAAGTAA	4805
Qy	4902	atgatgccaatataaagaacaggcatcaatgaagacaaagcctttgttaaaagacttgaaa	4961
Db	4806	ATGATGCCAATAAAAAGCAAGGCATCAATGAAGACAACGCCCTTTGTTAAAGGACTTTGAA	4865
Qy	4962	aagccgctcttgatacaaaaacaaacacgcgcagtaactgtgggtgatttaaatgcgcg	5021
Db	4866	ANGCGGCTCTTGATAACAAACCAAAACCGCGCAGTAAGTCTGGGTGATTTAAATGGCG	4925
Qy	5022	ttgcccaaacacgcctgaacctttgcaggggatcacaggacacacgcgtataaaaactgggcg	5081
Db	4926	TTGCCCAACACCGCTGACCTTTTCAGGGGATACAGGCACACAGGCTAAAAAACTGGGCG	4985
Qy	5082	agacttgacctatcaagggtgggcaacacagacaccaataagctaacccgataataacatcg	5141
Db	4986	AGACTTTGACCATCAAAAGTGGGCAACACAGACACCAATAAAGCTAAACCGATATAACATCG	5045
Qy	5142	gtgtggtagcagttactgatggcttcactgtcaaaacttgccaagaaccttaacaaatotta	5201
Db	5046	GTGTGGTAGCAGGTACTGATGGGTTTCACCTGTCAAACTGCCAAAGACCTTAACCAATCTTA	5105
Qy	5202	acagcgttaatgcagggtggcaccaaaatgatgacaaaaggcgtgtctttgttagactcaa	5261
Db	5106	ACAGGGTTAATGCAGGTGGTACCANAATTTGATGACAAAAGCGGTCTTTTGTAGACTCAA	5165
Qy	5262	gcggtcaagccaaagcaaacacccctgtcgtgaagtgcgaatggcctgggtggcgca	5321
Db	5166	GCGGTCGAAGCCAAAGCAAAACACCCCTGTGCTAAGTGCCAATGGGCTGGACCTGGGTGCA	5225
Qy	5322	aggtcatcagtaatgtgggcaaaaggcacaaaagataccgagcgtgccaatgtacaacagt	5381
Db	5226	AGGTCTATCAGCAATGTGGGCAAAAGGCACAAAAGACACCGACGCTGCCAATGTACACAGT	5285
Qy	5382	taaacgaagtacgcaacttgttgggtcttggttaatgtctggttaataacagctcagcgca	5441
Db	5286	TAAACGAAGTACCGCAACTTGTGGGTCTTTGGTAAATGCTGTTAATGATACGCTGACGGCA	5345
Qy	5442	atcaggtataacattggccgacatcaaaaaagaccaaattcaggtttcatctactaacgcga	5501


```
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-913-942-14

Query Match 1.2% Score 83.6; DB 4; Length 2037;
Best Local Similarity 54.1%; Pred. No. 2e-13;
Matches 198; Conservative 0; Mismatches 159; Indels 9; Gaps 1;

QY 6324 aaaaatggcagcaggtgaggtcagtgccaccagcaccgagtcggtcaatggtgacaggt 6383
Db 1670 AAAACGTCGGCGGAGGCGAAATTTCTGCCACTTCCACCAGATCGGATTAACGGAAAGCCAGT 1729
QY 6384 tgtac-----aaagccaccacccaagcatgtccaaacgcaaccaatgagcttgaccatc 6434
Db 1730 TGTATGTCGGTGGCAAAAGGGTAAACAAACCTTGCTGACAAAGTGAATTAATCTTGAGGGCA 1789
QY 6435 gtatccaccacaaacgaaataaaggccaatgacggaggtattcatcagcagatggcgatggcggt 6494
Db 1790 AAGTGAATAAAGTGGGCAACGTCAGATGCAGGTACTGCAAGTGCATTAGCGGCTTCAC 1849
QY 6495 ccattgccaaagcctacattcctggcagatccattggttaccgggttattgccaccaca 6554
Db 1850 AGTTACCACAGCCACTATGCCAGGTAAATCAATGTTCTATTGCGGGAAGTAGTTATC 1909
QY 6555 acgttcaaggtgaggtggcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgat 6614
Db 1910 AAGTCAAAATGTTTACGTATCGGGGTATCAAGAAATTCGATTAATGGCAAGTAGTTA 1969
QY 6615 ttaaaatcaatggttcagcagatatacccaagccatgtaggggggaggttgggtgcaggtt 6674
Db 1970 TTCCTTGTCTGGCACAACCAATAGTCAAGTAAACAGCGGTTGCAGCAGGTGTTGGTT 2029
QY 6675 tteact 6680
Db 2030 ACCAGT 2035

RESULT 4
US-08-409-995-1
; Sequence 1, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RTF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
US-08-409-995-1

Query Match 1.2% Score 82.8; DB 1; Length 3294;
Best Local Similarity 49.8%; Pred. No. 4.6e-13;
Matches 244; Conservative 0; Mismatches 237; Indels 9; Gaps 1;

QY 6200 aaatctgacggcagcaggttacacacccacagcaggtgcaaccgggtacaggtttaaaggc 6259
Db 2803 AAAGCTGACGGTACTCGGATAAAACCAAGGGAAGTGAGCAATGATAAAGTTTCTTACC 2862
QY 6260 ttgtctggacaaaacgggttggtggttcctcggtggtggtggtggtggtggtggtggtggt 6319
Db 2863 GATGAAAAACACGTTGTGAGCCTTTGATCCAAATGATCAATCAAAAGGTAAAGGTGTCGTG 2922
QY 6320 atccaaatgtggcagcaggtgaggtcagtgccaccagcaccgagtcggttcaatggtagc 6379
Db 2923 ATTGACAATGTGGGTAAATGGCGATATTTCTGCCACTTCCACCGATGCGATTAAACGGAAGT 2982
QY 6380 cagttgtac-----aaagccaccacccaagcattgccaaacgcaaccaatgagcttgac 6430
Db 2983 CAGTTGTATGCTGTGGCAAAAGGGTAAACAAACCTTGCTGACAAAGTGAATTAATCTTGAG 3042
QY 6431 catggtatccaccacaaacgaaataaaggccaatgacggaggtattcatcagcagatggcgatg 6490
Db 3043 GGCAAAAGTGAATAAAGTGGCAACCGTGCAGATGCAGGTACAGCAAGTGCATTAGCGGCT 3102
QY 6491 gcgtccatgcccagcctacattcctggcagatccatggttaccgggttattgccacc 6550
Db 3103 TCACAGTTACCACCAAGCCACTATGCCAGTAAATCAATGTTGCTATTGCGGGGAAGTAGT 3162
QY 6551 cacaacggtcaaggtgaggtggcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 6610
Db 3163 TATCAAGGTCAAAATGTTTACGTATCGGGGTATCAAGAAATTTCCGATAATGGCAAGTG 3222
QY 6611 gattttaaataatggttcagcagcagatatacccaagccatgtagggggcgagtggtgca 6670
Db 3223 ATTATTCGCTTGTGAGGCACAAACCAATAGTCAAGGTAAACAGCGGTTGCACAGGTTGT 3282
QY 6671 ggttttcaact 6680
Db 3283 GGTACCAGT 3292

RESULT 5
US-08-685-467-1
; Sequence 1, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
```



```

QY   4431 ccaccacccgctaaggtgacc 4450
      || | | | | | | |
DB   2741 CAGGTACAACACGGTAACC 2760

RESULT    9
US-08-685-467-3
; Sequence 3, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-685-467-3

Query Match          0.9%; Score 64.8; DB 3; Length 5738;
Best Local Similarity 66.4%; Pred. No. 8.6e-08;
Matches           93; Conservative     0; Mismatches 47; Indels     0; Gaps       0;

QY   4311 acaaaacccgtgcgcagcaatbtgtgatgcttaagcgacggctttaacttggaaggcca 4370
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   2621 AAAATCCAAATGCGACGAAGTATTGAAGATGTTATTCGCCGCAGGTGGGAATAATCAAGGTA 2680
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY   4371 atggtgaacggtttgaccttctcccatttatcacaccgtcaacctttgccgatggcaatg 4430
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   2681 ATGCTAATAATGTTTGNTATATGACGGCAGGTATGACACAGTAGTAACCTTACCAGTGNACACGA 2740
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY   4431 ccaccacccgctaaggtgacc 4450
      || | | | | | | |
DB   2741 CAGGTACAACACGGTAACC 2760


RESULT    10
us-09-377-155-1
; Sequence 1, Application us/09377155
; Patent No. 6197312
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 33595.7 Seconds
(without alignments)
3209.970 Million cell updates/sec

Title: us-09-361-619-5
Perfect score: 6972
Sequence: 1 ccatggtatgggcaggtgt.....ccgttgacctgcaggtgcac 6972

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vi: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_vil: *
59: gb_v12: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_rol: *
95: gb_rod: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	6972	100.0	6972	10	AX079917 Sequence
2	6961	99.8	6973	9	AR040716 Sequence
3	6961	99.8	6973	10	AX079913 Sequence
4	6258	89.8	6259	10	AX079924 Sequence
5	6141	88.1	6141	10	AX079918 Sequence
6	5976	85.7	5976	10	AX079914 Sequence
7	4671.2	67.0	62909	10	AX067457 Sequence
8	4101.6	58.8	6159	10	AX079920 Sequence


```

9      3135 45.0 3135 10 AX079957
10     2590 37.1 6942 10 AX079922
11     2448 35.1 2448 10 AX079959
12     426.8 6.1 1000 10 AX081547
13     94.2 1.4 10533 1 AE006194
14     84.6 1.2 3700 3 PM0277635
15     84.6 1.2 3700 3 HTU38617
16     84.4 1.2 11159 1 AE005587
17     84.4 1.2 28090 2 AP002566
18     82.8 1.2 3294 1 T55122
19     78.2 1.1 7062 3 U41852
20     76.8 1.1 11650 3 U32846
21     76.2 1.1 3324 3 STY277623
22     71.4 1.0 15021 1 AE003982
23     67.6 1.0 149789 60 AC007926
24     67.4 1.0 7921 5 AF153362
25     67 1.0 8139 3 PM0277636
26     67 1.0 10903 1 AE006108
27     66.6 1.0 10259 1 AE004017
28     64.8 0.9 5738 10 T55123
29     62.8 0.9 199551 60 AC006281
30     61 0.9 1993 1 AF125375
31     60.8 0.9 1602 5 AF036171
32     60.6 0.9 1518 6 DBAC11
33     60.4 0.9 11682 1 AE005230
34     60.4 0.9 262278 2 AP002552
35     60.2 0.9 1797 2 AF226381
36     60.2 0.9 1806 5 AF051898
37     60 0.9 60 9 AF040717
38     59.4 0.9 10965 1 AE002450
39     59.4 0.9 34980 9 AX044031
40     58.6 0.8 1770 2 AF226362
41     58.6 0.8 1770 2 AF226363
42     58.6 0.8 1770 2 AF226384
43     58.6 0.8 1797 2 AF226359
44     58 0.8 147838 76 AC083888
45     57.6 0.8 1779 9 AX020272

```

ALIGNMENTS

```

RESULT 1
AX079917 LOCUS AX079917 6972 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 5 from Patent WO0107619.
ACCESSION AX079917
VERSION AX079917.1 GI:13159438
KEYWORDS
SOURCE Moraxella catarrhalis.
          Moraxella catarrhalis.
          Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
          Moraxella.
REFERENCE 1 (bases 1 to 6972)
AUTHORS Loosmore, S.M., Sasaki, K., Yang, Y.P. and Klein, M.H.
TITLE Recombinant high molecular weight major outer membrane protein of
        moraxella
JOURNAL Patent: WO 0107619-A 5 01-FEB-2001;
        CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
        Location/Qualifiers
            source
                /organism="Moraxella catarrhalis"
                /db_xref="taxon:480"
BASE COUNT 2265 a 1555 c 1532 g 1620 t
ORIGIN
Query Match 100.0%; Score 6972; DB 10; Length 6972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ccatgatggcagtggtgctgcctccgctgatggcgatgacacccattggcc 60
    ||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db 1 CCATGGATATGGGCGAGTGCTGCTCGCCCTGCGGTATGATGGGGATGACACCCCATTTGCC 60
61 catactgtacgattgacatgtgatattgaacatgtgacatgtattgaacattgtt 120
    ||||||||||||||||||||||||||||||||||||||||||||||||
61 CATATCTGACGATTTGACATGTGATGATTTAAACATGTGACATGATTTAAACATGTTT 120
    ||||||||||||||||||||||||||||||||||||||||||||||||
121 aatctgttgcacatcaatacaataatttagtaacgcatttagtaacgcatttgaataa 180
    ||||||||||||||||||||||||||||||||||||||||||||||||
121 AATACTGTGTCATCATTAACATTAATTTAGTAAACGATTTAGTAAACGATTTGTAANA 180
    ||||||||||||||||||||||||||||||||||||||||||||||||
181 cattgcacccctttatgtatcatatgaataagaattattgattgattctgattattgt 240
    ||||||||||||||||||||||||||||||||||||||||||||||||
181 CATTGCGCCCTTTATGTATCATATGAATAGAATAATATGATTTGATTCGATGATTTGT 240
    ||||||||||||||||||||||||||||||||||||||||||||||||
241 atcagaatggtagctatatgatgatgcctacagattgattgggttaactcaactctatg 300
    ||||||||||||||||||||||||||||||||||||||||||||||||
241 ATCAGAATGGTAGCTATATGATGATGCTACGCTACGAGTTGATTTGGGTTAATCACTCT 300
    ||||||||||||||||||||||||||||||||||||||||||||||||
301 atttgatatatttgaactaatactattgacttaataacacacacacacacacacacac 360
    ||||||||||||||||||||||||||||||||||||||||||||||||
301 ATTTGATATATTTTGAACCTAATCTATTGACTTAATCAACCATATGTTTAAATTTAGCA 360
    ||||||||||||||||||||||||||||||||||||||||||||||||
361 taatggtaggcttttgaataatcacatcgcaatattgttctactgttactaccatgct 420
    ||||||||||||||||||||||||||||||||||||||||||||||||
361 TAAATGGTAGGCTTTTGTAAATAATCACTCGCAATATTTGTTCTACTGTTACTACCATG 420
    ||||||||||||||||||||||||||||||||||||||||||||||||
421 tgaatgacgatcccaatccacagattcattcaagtgtgattgttatacgcacacattta 480
    ||||||||||||||||||||||||||||||||||||||||||||||||
421 TGAATGACGATCCCAATCACCAGATTTCATTCAAGTGATGTTTGTATACGACGACATT 480
    ||||||||||||||||||||||||||||||||||||||||||||||||
481 ccctaattattcaatacaaatgcctatgtcagatgtatcatttttttaaggttaaacac 540
    ||||||||||||||||||||||||||||||||||||||||||||||||
481 CCCTAATTTTTCATCAATCAATGCTTGTGACGATGATATCAATTTTAAAGGTAAACAC 540
    ||||||||||||||||||||||||||||||||||||||||||||||||
541 catgaatcacatctataaagtcatcttttaacaaagccacagcacatttattgagcagtg 600
    ||||||||||||||||||||||||||||||||||||||||||||||||
541 CATGAATCACATCTATAAGTTCATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGC 600
    ||||||||||||||||||||||||||||||||||||||||||||||||
601 agagtacgcaaaaatccacacagcagcgggggggggggggggggggggggggggggggg 660
    ||||||||||||||||||||||||||||||||||||||||||||||||
601 AGAGTACGCCAAATCCCAACAGACGCGGGGGGGGGTGTGCTGTCTACAGGGGCAAGTGG 660
    ||||||||||||||||||||||||||||||||||||||||||||||||
661 tctgcaactgtgagcttgcctgattgctgcgcgctgctgctgctgctgctgctgctgctg 720
    ||||||||||||||||||||||||||||||||||||||||||||||||
661 TGTATGCACTCTGAGCTTTGCCCGGTATTTGCCGCGCTGCTGCTGCTGCTGCTGCTGCTG 720
    ||||||||||||||||||||||||||||||||||||||||||||||||
721 gctcagtgggcagtgcttatgctcaaaaaaagataaccaaacatcgcacattcgggtgaa 780
    ||||||||||||||||||||||||||||||||||||||||||||||||
721 GCTCAGTGGCAGTGCTTATGCTCAAAAAAAGATACCAAAACATATCGCAATTTGGTGAAC 780
    ||||||||||||||||||||||||||||||||||||||||||||||||
781 aaaccagccaagacgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
    ||||||||||||||||||||||||||||||||||||||||||||||||
781 AAACAGCAAGACGCTCAGGCACCTGCCAAGCGGACGCGGTGATCGAGCCATTGCTATTGG 840
    ||||||||||||||||||||||||||||||||||||||||||||||||
841 tgaataatcgaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900
    ||||||||||||||||||||||||||||||||||||||||||||||||
841 TGAATAATGCTTACGCACAGGGGGGTCAAGCCATCGCCATCGGTAGTAGTAGTATATAAAC 900
    ||||||||||||||||||||||||||||||||||||||||||||||||
901 caatggaagcagcttgggaagaataggtagccagatgctcagcagcagcagcagcagcagc 960
    ||||||||||||||||||||||||||||||||||||||||||||||||
901 CAATGGAAGCAGCTTTGGGATAAGATAGGTACCGATGCTACGGGTCAAGAGTCCATCGCC 960
    ||||||||||||||||||||||||||||||||||||||||||||||||
961 cgggtggtgatgtaaaaggctagtggtgatgcctgattgctcagcagcagcagcagcagcag 1020
    ||||||||||||||||||||||||||||||||||||||||||||||||
961 CGGTGGTGATGTAAGGGCTAGTGGTGATGCGCTCGATTGCCATCGGTAGTAGTAGCTTACA 1020
    ||||||||||||||||||||||||||||||||||||||||||||||||
1021 ttgtgttgatcagcatgggtaatacctaataacacacacacacacacacacacacacac 1080
    ||||||||||||||||||||||||||||||||||||||||||||||||
1021 TTTGCTTGATCAGCATGGTAACTCTTAACATCCGAAAGGTACTCTGTGATTAACGATCTTA 1080
    ||||||||||||||||||||||||||||||||||||||||||||||||
1081 taacggccatgcagctattaaagaataacgaagctcgaagcagcagcagcagcagcagcag 1140
    ||||||||||||||||||||||||||||||||||||||||||||||||
1081 TAACGCCCATGCGAGTATTAAAGAAATACGAAGCTCAAGAGGATAATGATGTAAAAATATAG 1140
    ||||||||||||||||||||||||||||||||||||||||||||||||

```


QY 1141 acgcacaaaccgcaagggacacccagctactgcagtgaggagccatgtcatatgcacaggg 1200
 Db 1141 ACGCACAACCGCAAGCGGACACGCCAGTACTGCGAGTGGGAGCCATGTCTATGCACAGGG 1200
 QY 1201 tcatttttcaacgccttttggttacagggcaacagctaaagtgccctattccctggcagt 1260
 Db 1201 TCATTTTTCAACGCCCTTTGGTACACGGGCAACAGCTTAAAGTGCCTATTTCCTTGGCAGT 1260
 QY 1261 gggcttgcgcacagcgagggccaaactacaactcgctattggcttctgatcacaacatc 1320
 Db 1261 GGCTCTGGCGCCACAGCGGAGGCCAATCTACAATCGCTATTGGTCTGATGCACATC 1320
 QY 1321 tagctcgttggggcgatagcccttgggtgcaggtactcgtcgcagctacagggcgagt 1380
 Db 1321 TAGCTCGTTGGGAGCATAGCCCTTGGTGCAGGTACTCGTGCCTCAGCTACAGGCGCAGTAT 1380
 QY 1381 tgccttaggtcaaggtctgttctactcagagtgatgataataattctagaccggcctatc 1440
 Db 1381 TGCCCTTAGGTCAGGTTCTGTTGTCTACTCAGAGTGATAATAATCTAGACCGGCCCTATAC 1440
 QY 1441 accaaaTaccagggcactagaccccaagtttcaagccaccaataaTaccagagcggggtcc 1500
 Db 1441 ACCAAATACCCAGGCACACTAGACCCCAAGTTTCAAGCCACCACAAATAATACAGAGCGGGTCC 1500
 QY 1501 actttccattggttagtaactctatcaaaogtaaaatcaatcgtcgggtcaggtgttaa 1560
 Db 1501 ACTTTCATTGGTAGTAACCTCTATCAACCGTAAATCATCAATGTCGGTGCAGGTGTAA 1560
 QY 1561 taaaaccgatcggtcaattggtgcacagctagaagcgtgtgtaagtgggtcgaagagcg 1620
 Db 1561 TAAACCGATGCGGTCGAATGTGCACAGCTAGAGCGGTGGTGAAGTGGGCTAAGGAGCG 1620
 QY 1621 tagaattacttttcaggggtgatataacagctactgcgttaaaataggtttgtgataatc 1680
 Db 1621 TAGAATTACTTTTCAGGGTGATGATTAACAGTACTGACGTAAAAATAGGTTTGGATAATAC 1680
 QY 1681 tttaactattaaaggttggtgcagagaccacgcattaaacgcgtataataatcgtgtggt 1740
 Db 1681 TTTAACTATTAAAGGTGGTGCAGAGACCAACGCATTAAACGATAAATAATTCGGTGTGGT 1740
 QY 1741 aaagagctgataatagtggtctgaaagttaaactgtctaaactttaacaatttacc 1800
 Db 1741 AAAGAGGCTGATAAATAGTGGTCTGAAAGTTAACTTTGCTTAAACCTTTTAAACAATCTTAC 1800
 QY 1801 tgagggtaaTacaactacattaaatgcacaacacagttaaaggttagttagtagtag 1860
 Db 1801 TGAGGTGAATACAACATACATTAAATGCCACACACAGTTAAAGGTAGGTAGTAGTAG 1860
 QY 1861 tactacagctgaattattgagtgatagtttaacctttaccgcccaatacagggcagtc 1920
 Db 1861 TACTACAGCTGGAATTATTGAGTGATAGTTTAAGCTTTTACCAGCCCAATAACAGCGAGTCA 1920
 QY 1921 aagcacaagcaaaacccgtctatgcttaagtggtgaaagttacttaataatcagaac 1980
 Db 1921 AACCAAGCAAAACCCGCTATGGCGTTAATGGGGTGAAGTTTACTTAATAATGCAGAAAC 1980
 QY 1981 aacagcagcaatcggcactactcgtattaccagagaTaaaattggctttgctcgagatgg 2040
 Db 1981 AACAGCAGCAATCGGCACACTACTCGTATTACCAGAGATAAAATTTGGCTTTCCTCGAGATGG 2040
 QY 2041 tgatgttgaTgaaaaaaagcaccatatttggataaaaaaacaacttaagtggttagtgt 2100
 Db 2041 TGATGTGTGATGAATAAACAGCAGATATTGGGATAAAAAACAACCTTAAAGTGGGTAGTGT 2100
 QY 2101 tgaattaccatagacaatggcattgatgcagtaataataaagatcagtaattctgcaa 2160
 Db 2101 TGAATTTTACCATAGACAATGGCATTTGATGCAGGTAAATAAAAAAGATAGTAATCTTGCCAA 2160
 QY 2161 aggtagcagtgctaaagcagtcggttaccatcgacagctcaagcgccgaagcactcttt 2220
 Db 2161 AGGTAGCAGTGTCAACGATCGGTTACCATCGAACAGCTCAAGCGCCCAAGCCCTACTTTT 2220

QY 2221 aaacgcagggcgctggcgtcatcagtgTcacaccctactgaaatatcagtgctgatgaagatgg 2280
 Db 2221 AAACGCAGGGCGCTGGCATCAGTGTACACCCCTACTACTGAATAATCAGTTGATGTAAGAGTGG 2280
 QY 2281 caatgttacgcgcaccaacttacaacattggcgtgaaacaccagcagcttaacagtgatgg 2340
 Db 2281 CAATGTTACGCCCCCACTTTACAACATTGGCGTGAACACCACCGAGCTTAAACAGTGATGG 2340
 QY 2341 cactagtataaaatttagttgtaagggttagtggTaccgaaacaataagcttagttaccgcca 2400
 Db 2341 CACTAGTGATAAATTTAGTGTTAAGGCTAGTGGTACGAACAATAGCTTAGTTACCGCCGA 2400
 QY 2401 acatttggcaagctatctaaatgaagtcaatcgaacggctgacagtgctctacaagaact 2460
 Db 2401 ACATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGGCTGACAGTGTCTCTACAAAGCTT 2460
 QY 2461 tacggttaagaagaacacatgatgcgcacaacgcttatcacggtgctaaagatacagac 2520
 Db 2461 TACCGTTTAAAGAAGAACGATGATGACGCCACGCTATCACCGTGGCTAAAGATACGAC 2520
 QY 2521 aaaaaaTgcgggcgcagtcagcatcttaaaactcaaaaggttaaaacggtctaaacggttgc 2580
 Db 2521 AAAAAATGCCGGCGCAGTCAGCATCTTAAACTCAAAGGTAAACACGGTCTAACGGTTGC 2580
 QY 2581 taaaaaaagatgggttacggttacctttgggttagccaagatagcggtctgcaccattgg 2640
 Db 2581 TACCAAAAAAGATGGTAGCGTTACCTTTGGGCTTAGCCAAGATAGCGGTCTGACCATTGG 2640
 QY 2641 caaaagcaccctaaacaacgatggcttgactgtttaagataccaaacgaacaaatccaagt 2700
 Db 2641 CAAAAGCACCCTTAAACAACGATGGCTTGACTGTGTTAAAGATACCAACGAACAAATTCCAAGT 2700
 QY 2701 cgggtcctaaggcattaaatttactaatgtgaatggTtagtaatccaggtactcgtgcatgg 2760
 Db 2701 CGGTGCTAATGGCATTAAATTTACTAATGTGAATGGTAGTAATCCAGGTACTGGCATGGC 2760
 QY 2761 aaatccgcctgcgaattaccagagataaaattggcttgcgttgcgtctgatggtgcagttga 2820
 Db 2761 AAATACCGCTCGCATTACCAGAGATAAAATTTGGCTTTGCTGTTCTGATGTGTCAGTTGA 2820
 QY 2821 taaaaaacccttctctgaatcaagaacagctacaagttggcaattgaagataccaa 2880
 Db 2821 TACAAACAACCTTATCTTATCAAGCAAGCTACAAGTTGGCAATGTTAAGATTACCAA 2880
 QY 2881 cactggcattaaacgcaggtgtgtaagccatcacaggcgtgtcccccaacactgcctagcat 2940
 Db 2881 CACTGGCATTAAACGCAGTGTGTAAAGCCATCACAGGGCTGTCCCCACACACTGCCTAGCAT 2940
 QY 2941 tgccgatcaaaagtagcgcgaacataagaactgggcaatatacaatccaagacaagaacaatc 3000
 Db 2941 TGCCGATCAAAAGTAGCCGCAACATAGAACTGGGCAATACAAATCCCAAGACAAGACAATC 3000
 QY 3001 caacgctgccagcattaatgatatattaaatcacaggctttaaccctaaaaataatacaaa 3060
 Db 3001 CAACGCTGCCAGCATTAATGATATATTAAATACAGGCTTTAAACCTTAAAAATAATAACAA 3060
 QY 3061 ccccatgactttgtctccactttagcattgttgactttgccaatggccaatggcaccac 3120
 Db 3061 CCCCATTGACTTTGTCTCCACTTATGACATTGTTGACTTTGCCAATGGCAATGCCACCAC 3120
 QY 3121 cgcaacagtaaccatgataccgcttaacaaaaccagtaaaagtggtatgatgtgaatgt 3180
 Db 3121 CGCCACAGTAGTAACCCATGATACCGCTAAACAAACCCAGTAAAGTGTATATGATGTGAATGT 3180
 QY 3181 ggatgatcaaacattcatctaacagcactgatgacaataaaacacttggcgtcaaaaac 3240
 Db 3181 GGATGATACAAACCTTATCTTAACAGGCACCTGATGACAATAAAAAACTTGGCGTCAAAAC 3240
 QY 3241 caccaaaactgaacaaaacaaagtgctaaTggttaatcacagcaactaactttaatgttaactc 3300
 Db 3241 CACCAAACTGAACAAAACAAAGTGTAAATGGTAAATACAGCAACTAACTTAAATGTAACTC 3300
 QY 3301 tagtgatgaagatgcccttgttaaagcgaagacatcgccgaaaaatctaaacacccctagc 3360

|||||
Db 3301 TAGTGATGAGATGCCCTTGTAAAGCCAAAGACATGCCGAAAATCTAAACACCCTAGC 3360
Qy 3361 caaggaaattcaaccaccacaaaggacagacagacacgcccctcaaaacctttacggttaa 3420
Db 3361 CAAGGAAATCACACCCACCAAGGACAGCAGACACCGCCCTCAAAACCTTTACCGTTAA 3420
Qy 3421 aaaggtagatgaaataataatgctgtagacgcaagcccatcacccgltgggtcaaaagaa 3480
Db 3421 AAAGGTAGATGAAATATATGCTGATGACGCCAACGCCATACCCGTGGGTCAAAGAA 3480
Qy 3481 cgcataataatcaagtcaaacaccccttaacactcaagggtgaagacgggtcttaataataaac 3540
Db 3481 CGCAATAATCAAGTCAACACCCCTAAACACTCAAAAGGTGAAACGGCTCTTAATATTAAAC 3540
Qy 3541 cgcataaattgtagcttaccttggcattacacacacacagcgggtcttaagccggcaaa 3600
Db 3541 CGCAATAATGGTAGGTTACCTTTGGCATTAAACACCAAGCGGCTTTAAAGCCGGCAA 3600
Qy 3601 aagcaccctaaacgacggtggtcttctattaaaaacccccactgtagcgaacaaatcca 3660
Db 3601 AAGCACCTTAACGACGGTGGCTTGTCTATTAAACCCCACTGGTAGCGAAACATCCA 3660
Qy 3661 agtcgggtcgtatgagtgaggttggcaaggttaataataatggttctgtagtgctgg 3720
Db 3661 AGTCGGTCTGATGGCGTGAAGTTTGGCAAGGTTAATAATATGTTGTTAGTGGCTGG 3720
Qy 3721 cattgtagcacaactgcattaccagagatgaattggctttactggactaatggctc 3780
Db 3721 CATTGATGGCAACACTGCCATTACAGAGATGAATTTGGCTTTACTGGGACTTAATGGCTC 3780
Qy 3781 acttgataaaagcaaacccccacctaagcaagagcgcaattaacgcaggtggttaaaagat 3840
Db 3781 ACTTGATAAAAGCAAAACCCCACTTAAGCAAAAGCGCATTAACGCAGGTGGTTAAAGAT 3840
Qy 3841 tacaacaattcaatcaggtaggattggcccaaaacagccatgactgctgtagcaggcgcaa 3900
Db 3841 TACCACCAATTCAATCAGGTGAGATTGCCAAACAGCCATGATGCTGTGACAGCGCGCAA 3900
Qy 3901 gatttatgatttaaaacccgaacttgaaacaaaatacagcagtagctgcacaaacagcaca 3960
Db 3901 GATTTATGATTTAAACCCGAACTTGAAACAAANAATCAGCAGTACTGCCAANAACGACACA 3960
Qy 3961 aaactcattacagaaattctcagtagcagatgaacaaaggttaataactttacggttagtaa 4020
Db 3961 AAACCTCATTTACAGCAATTTCTCAGTAGCAGATGAACAAGGTAAATAACTTTACGGTTAGTAA 4020
Qy 4021 cccctactccagttatcacacctcaaaacotctgattcactaccccttgcagatgaaa 4080
Db 4021 CCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCATCACCCTTTGCAGGTTGAAA 4080
Qy 4081 cggcattaccacccagggttaataaagggtggtgctggtggcattgaccacaaacaaagg 4140
Db 4081 CGGCATTACCACCAAGGTAAATAAGGTGTGGTGGTGGCATTTGACCAACCAAGG 4140
Qy 4141 cttaaccacgctaaagtgcacgctgggtggttaataataatggcaaaagcattgtcaattgacag 4200
Db 4141 CTTAACCAACCGCTTAAGCTGACCGTGGGTAAATAATATGGCAAGGCAATTTGATTGACAG 4200
Qy 4201 ccaaaatggtcaaaataccatcacaggactaagcaaacacttagtcaatglttaccatga 4260
Db 4201 CCAAAATGGTCAANAATACCTACAGGACTAAGCNACACTCTAGCTAATGTTTACCNAATGA 4260
Qy 4261 taaagggtgctgcacacacagaggggcaataataataaagacgaagcaaaacccg 4320
Db 4261 TAAAGGTAGCTAGCCACACAGAACAGGGCAATATAATCAAGACGAAGACAAACCCG 4320
Qy 4321 tgcgcacagcattgtgtagtgctaaagcagcaggttttaacttgaaggaatggtgaagc 4380
Db 4321 TGCCGCCACGCAATTTGATGTGCTTAAGCCCAAGGCTTTAACTTGCAGGCAATGTGTGAAGC 4380
Qy 4381 ggttgacttgtctccacttatgacacccgtcaactttgccgatggcaatgcccaccgcg 4440
|||||

Db 4381 GGTGACTTTTGTCTCCACTTTATGACACCGTCACACTTTGCCGATGGCAATGCCACACCGC 4440
Qy 4441 taagtgaccttatgatgacacaagcaaaacagtaaaagtgcttatgatgtcaatgtgga 4500
Db 4441 TAAAGTGCACCTTATGATGACACAAAGCAAAACCAAGTAAGTGGTCTATGATGTCAATGTGGA 4500
Qy 4501 tgatataccacttgaagttaaaagataaaaaaacttggcgtaaaaaacccacacattgaccag 4560
Db 4501 TGATACAAACCAATTCAAGTTAAAGATAAAAAACTTGGCCTTAAAAACCAACCACTTTGACCAG 4560
Qy 4561 tactggcaaggtgctctaataatttggcctaagaacaataaagctactggcgatggcctgt 4620
Db 4561 TACTGGCACAGGTCTAATAAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGGCGCTTGT 4620
Qy 4621 caagccagtgatgctgctcatcaataaacccttactgtgcacatcccaaacctgcaaa 4680
Db 4621 CAAGGCCAGTATATGCTTGTCTATCTAAACACCTTATCTGGGACATCCAAACTGCCAA 4680
Qy 4681 aggggcaagccaaagcaaacactcagcaggttatgtgtagtctgtagtgcataaaggtcat 4740
Db 4681 AGGGCAAGCCAAAGCGAAACAACCTCAGCAGGCTATGTGGATGCTGATGGCAATTAAGTCTAT 4740
Qy 4741 ctatgacagtagccgataaacaagtactatcaagccaaaataatgatggcacagttgataaac 4800
Db 4741 CTATGACAGTACCGGATAACAAGTACTATCAAGCCAAAAAATGATGGCACAGTTGATATAAC 4800
Qy 4801 caaagaagttgccaaagcaaacctggtgcgcaagcccaaacccccagatggcacattggc 4860
Db 4801 CAAGAAGTTGCCAAAGACAAACTGGTGCCTCAAGCCCAACCCAGATGGCACATTGGC 4860
Qy 4861 tcaaatgaatgtcaaatcagtcatttaacaaagaacaaagtaaatgatgccaataaaaaagca 4920
Db 4861 TCAATGAATGTCAAAATCAGTCATTACAAAAGACAAGTAATGATGTCATAATAAAGCA 4920
Qy 4921 aggcataataaagacacgccttgttaagagccttgaagagcccttctctgtataacaa 4980
Db 4921 AGGCATCAATGAAGACAACGGCTTTGTTAAAGGACTTGAANAAGCCGCTTCTGTATAAACA 4980
Qy 4981 aaccaaaccgcgcagtaactggtggtgatttaaatgctgtgcccaaacacgcctgac 5040
Db 4981 AACCAAAAACGCCAGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCCCTGAC 5040
Qy 5041 ctttgcagggtatcacaggacacaacgcttaaaaaactggcgagaccttggaccatcaaaagg 5100
Db 5041 CTTTGCAGGGATACAGCACACCGCTTAAAAAATGGCGGAGACTTTTACCATCAAAAGG 5100
Qy 5101 tggcaaacagacacccaataagctaaaccgataataaacatcggtggtgtagcaggtactga 5160
Db 5101 TGGGCAACACACACCAATAGCTAACCGATATAAACATCGGTGGTGGTAGCAGGTACTGA 5160
Qy 5161 tggcttcaactgtcaaaacttgcacaaagacctaaccatcttcaacagcgttaatgcagggtgg 5220
Db 5161 TGGCTTCACTGTCAAACTTGCCAAAGACCTAACCAATCTTAACAGCGTTAATGTCAGGTGG 5220
Qy 5221 caccataatgatgacaaaaggcgtctctttgttagactcaagcgtcaagccaaagcaaa 5280
Db 5221 CACCAAAATGATGACAAAGCGCTGCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAAA 5280
Qy 5281 caccctgtgctaaagtgcacaaatgggctggacccgtgggtggcaagggtcatcagtaatgtggg 5340
Db 5281 CACCCCTGTCTAAGTGCCATGGCTGGACCTGGTGGTGGCAAGGTCATCAGTAATGTGGG 5340
Qy 5341 caaaggcacaagaataccgacgctgccaatgtcaaacagtttaaacgaagtacgcaactt 5400
Db 5341 CAAAGGCCAAAGATACCGCGCTGCCAATGTACAAGCTTAAACGAAGTACGCAACTT 5400
Qy 5401 gttgggtcttgtaatgctgtaataatgaataacgctacagcgaatcaggtataacattgcgca 5460
Db 5401 GTTGGGTCTTGTGTAATGCTGGTAATGAACCTCAGCGCAATCAGGTAAACATTGTCCGA 5460
Qy 5461 catcaaaaaagaccacaaatcagggttcattcttaacgcgactgtcatcaagcagggcac 5520
Db 5461 CATCAAAAAGACCCAAATTCAGGTTTCATCTTAACCGCACTGTCTATCAAAAGAGGCAC 5520

QY	6601	tgggtcaatgggtatttaaatacaatgggttcacgccgataccccaggccatgtagggcgccg	6666
Db	6601	TGGTCAATGGGTATTAAATCAATGGTTTCAGCCGATACCCCAAGGCATGTAGGGCGGC	6666
QY	6661	agttggtcaggttttcacttttaagccataaatcgaagattttacttataaatacaatc	6720
Db	6661	AGTTGGTCAGGTTTTCACTTTAAAGCCATAATCCGAAGATTTTACTTTAAATAACAATC	6720
QY	6721	tcaccatagttgtataaaacagcatcagcatcagtcagtcataattactgatgctgattgtttt	6780
Db	6721	TCACCATAGTTGTATAAAACAGCATCAGCATCAGTCATATTACTGATGCTGATGTTT	6780
QY	6781	atcaactaaaccattttacccgtccaagtattctcttccaacatgacccaataatcgccattg	6840
Db	6781	ATCACTTAAACCATTTTACCGCTCAAGTGATTCCTCTTCCACCATGACCAATCGCCATTG	6840
QY	6841	atcataggtaaacttattgaataaattttatacaatagtttgtatagatatgggttaaaatt	6900
Db	6841	ATCATAGGTAAACTTATTGAGTAATAATTTTATCAATGATGTTGTAGATATGGTTAAAT	6900
QY	6901	gtgccattgacccaaaaatgaccgatttaccgcgaataattccgcgaataattctgattatgatccgttgac	6960
Db	6901	GTGCCATTGACCAAAAAATGACCGATTATTCGCCGAAAAATTTCTGATTATGATCCGTTGAC	6960
QY	6961	ctgcagggtcgac	6972
Db	6961	CTGCAGGTGCGAC	6972
RESULT 2			
AR040716			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches: 6972; Conservative			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			

Db 241 ATCAGAATGGTGATGCTATATGATGATGCCGTACGAGTTGATTTGGGTTAATCACTCTATG 300
Qy 301 atttgatataatttgaactaatctattgactttaataacocatatggttataatttagca 360
Db 301 ATTGTGATATATTTTGAAACTAATCTATGTACTTAATACCATATGTTTATAATTTAGCA 360
Qy 361 taatgtaggttttttgaataatcaatocgcaatatttcttactgttactaccatgct 420
Db 361 TAATGGTAGGCTTTTGTGTAATAATCACATCGCAATATTTGTTCTACTGTACTACCATGCT 420
Qy 421 tgaatgagatcccaatcaccagattcaattcaagtgatgtgtttgtatagcaccattta 480
Db 421 TGAATGAGATCCCAATCACCAGATTCAATTCAGTGATGTTTGTATACGCACCAATTA 480
Qy 481 ccctaattatttcaataatgccttatgtcagcatgtatcatcttttttaaggttaaacac 540
Db 481 CCCTAATTAATTTCAATCAAAATGCCATATCTCAGCATGTATCATATTTTAAAGGTAAACAC 540
Qy 541 catgaatcaatctataaagtcattctttaacaaagccacagcacatttatggcagtgcc 600
Db 541 CATGAATCACATCTATAAGTCAATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGCC 600
Qy 601 aagtagcgcgaatccacagcac-ggggggggtagctgtgtctacagggcaagtggca 659
Db 601 AGAGTAGCCCAATCCACAGCAGCGGGGGGGGTAGCTGTCTACAGGCAAGTTGGCA 660
Qy 660 gtgtatgactctgagctttgcccgtattgcccgtctgcgctgcgtctcctgtatcggtgcaa 719
Db 661 GTGTATGCACTCTGAGCTTTGCGCGTATTTGCGCGCTCGCTCTCTCGTGATCGGTGCAA 720
Qy 720 cgtcagtgagctgtctatctcaaaaaaagataaccaaatatcgcgaattggtgaac 779
Db 721 CGCTAGTGGCAGTGTCTATGTCTCAAAAAAAGATACCAACATATCGCAATGTTGTAAC 780
Qy 780 aaacagcgaagcagctcagcactgcgaagcggaagcggtgatcgagccattgctattg 839
Db 781 AAAACAGCCACAGCCTCAGGCATTCGCAAGCGGACGGTATCGAGCCATTTGATATTG 840
Qy 840 gtgaataatgctaacgacagcggtgcagccatcgccatcggttagtagtaataaaactg 899
Db 841 GTGAAATGCTTAACGACAGCGGCGGTCAAGCCATCGCCATCGGTAGTACTAATAAACTG 900
Qy 900 tcaatggaagcagctttggtaagataggttacogattgctacgggtcaagagtcacatgcga 959
Db 901 TCAATGGAAGCAGTTTGGATAAGATAGTACCGATGCTACGGGTCAAGAGTCCATCGCCA 960
Qy 960 tcggtgtgtataaagctagtggtgatccctcgattgccaatcggttagtgacttac 1019
Db 961 TCGTGTGTATGTAAAGGCTAGTGGTGTGTCCTCGATTGCCATCGGTAGTATGACTTAC 1020
Qy 1020 atttgcttgatcagcatggttaactcttaaacatccgaaggtactctgtattaacgatctta 1079
Db 1021 ATTTTGCTTGATCAGCATGGTAATCTTAACATCCGAAGGTACTCTGATTTAAGCATTTA 1080
Qy 1080 ttaacgccaatgagtatataaagaaaaatacgaagctcaaaaggaataatgataaataata 1139
Db 1081 TTAACGGCCATGCAAGTATTTAAAGAAATACGAAGCTCAAAGGATAATGATCTAAATATA 1140
Qy 1140 gacgcacaacccgcaagcggaacagccagctactgagtgaggccaatgtcatatgcaag 1199
Db 1141 GACGCACAACCGCAAGCGGACACGCCAGTACTGCAAGTGGGAGGCCATGTCATATGCACAG 1200
Qy 1200 gtcatatttccaacgcttttggtacagggcaacagataaaagtcattctcttgagcag 1259
Db 1201 GTCAATTTTCCAAACGCCCTTTGGTACACGGGCAACAGCTAAAGTGGCCATTTCTTGGCAG 1260
Qy 1260 lgggtcttgccgcacagcaggggcaaatctacaatcgctattggttcttgatgcaaat 1319
Db 1261 TGGGTCTTTGCCGCCACACCGGAGGCCAATCTACAATCGCTATTGTTGTTCTGATGCAACAT 1320
Qy 1320 ctatgctgtgggagcgtatgccccttggtgaggtactcgtcagctacagggcagta 1379
Db 1321 CTAGCTCGTTGGAGCGATAGCCCTTGGTGCAGGTACTCGTGCAGCTACAGGGCAGTA 1380

Qy 1380 ttgcccttaggtcaaggttctgttgcactcagagtgaataatcttagacgcgcctata 1439
Db 1381 TTGCCCTAGGTCAAGTCTGTGTGTCTACAGTATATAATTAATCTAGACCGGCTATA 1440
Qy 1440 caccaaatccccaggcactagaccocccaagtttcaagcccaacaataatacgaagcggggctc 1499
Db 1441 CACCAAAATCCCAAGGCACGTAGACCCCAAGTTTCAAGCCCAATATATAGNAAGGGGGTCT 1500
Qy 1500 caatttccattggtatgaactctatcaaaagtaaaatcatcaatgctcggtgcaggtgtta 1559
Db 1501 CACTTTCCATTTGTTAGTAACTCTATCAACGTAAATCATCAATGTCGCTGCAGGTTGTTA 1560
Qy 1560 ataaaacccgatgcggtcaaatgtggcacagctagaagcggtggtgaagtgggctaaaggagc 1619
Db 1561 ATAAACCGATGCGTCAATGTGGCAGCTAGAACGGTGTGTGAAGTGGCTTAAGGAGC 1620
Qy 1620 gtgaattacttttccagggttgatgaataacagtaactgaactgaataaaataggtttgataata 1679
Db 1621 GTAGAATTACTTTTCAAGGGTGATGAACAGTACTGACGTAAATAATAGGTTTGGATANTA 1680
Qy 1680 ctttaactattaaaggtggtgcagagaccacgcatttaaccgataataatcctggtgtg 1739
Db 1681 CTTTAACTATTAAAGTGTGTGCAGAGACCAACGCATTTAACCGATAATAATATCGGTGTG 1740
Qy 1740 taaaagaggctgataatagtggtggtgaaagttaaaacttgctaaacacacacacacacac 1799
Db 1741 TAAAGAGGGCTGATAATAGTGGTCTGAAGTTTAAACTTGTAAACCTTTAAACAATCTTA 1800
Qy 1800 ctgaggtgaataacactacatactaaatgcccaaccacagttaaggttaggtagtagtagta 1859
Db 1801 CTGAGGTGAATACAACTACATTAANTGCCACACACACAGTTAAGTATAGTATAGTA 1860
Qy 1860 gtactcagctgaattattgtagtgatgttttaaccccttaccagcccaacacacacacacac 1919
Db 1861 GFACTACAGCTGAATTTATGAGTGATAGTTTAACTTTTACCAGCCCAATACAGGCGATC 1920
Qy 1920 aaagcacaagcaaaacgcgtctatgctgaatgggtgaaagttaacttaataatgcagaaa 1979
Db 1921 AAAGCACAAAGCAAAACCGCTCTATGGCGTTAATGGGGTGAAGTTTACTTAATAATCA 1980
Qy 1980 caacagcagcaaatcgccactactcgtattaccagagataaaatggcttgcctcgaagtg 2039
Db 1981 CAACAGCAGCAATCGGCATCTACTCGTATTACAGAGATATAAATTTGGCTTTGCTCGAG 2040
Qy 2040 gtgactgtgataaaacaacagccacatattggataaaaaaacaacttaaaagtggtgagtg 2099
Db 2041 GTGATGTTGATGAAACCAAGCACCCATATTGGATTAATAAACCACTTAAAGTGGGTAGT 2100
Qy 2100 ttgcaattaccatagacaatggcattgatgcaggtaataaaaaagatcagtaactcttgcca 2159
Db 2101 TTGCAATTTACCATACACAAATGGCATTTGATGCAAGTAAATAAAAGATCAATCTTGCCA 2160
Qy 2160 aaggtagcagtgctaacgatgcggttacatcgaaacagctcaaaagccgcgcacactct 2219
Db 2161 AAGGTAGCAGTCTTAACGATGCGGTTTACCATTGAAACAGCTCAAAAGCCCAAGCCCTACT 2220
Qy 2220 taaagcagcgctggtgcatcaggtcacacccactgacacacacacacacacacacacacac 2279
Db 2221 TAAACGCAGCGCTGGCATCAGTGTGCACACCTACTGMAATATATCAGTTGATGCTAAGAG 2280
Qy 2280 gcaatgttaccgcccccaacttacaacattggcggtgaaacacccagcgttaacagtgatg 2339
Db 2281 GCAATGTTTACCGCCCCCACTTACAACATTTGGCGTGAACACCCAGCGCTTACACAGTATG 2340
Qy 2340 gcaatgtgataaaatttagtttaagggtagtgggtacgaacaataatagcttagtaccgcg 2399
Db 2341 GCACTAGTGTAAATTTAGTTTAAAGGTAGTGGTACGAACAATAAGCTTAGTTACCGCG 2400
Qy 2400 aacatttggcaagctctctaaatgaagtcgaacgcgctgacagtgctctacaaagct 2459
Db 2401 AACATTTGGCAAGCTATCTAAATGAATGAAGTCAATGCAAGCGCTGACAGTGTCTACAAAGCT 2460

QY 2460 ttaccgltaaagaagaagcagatgtagcgcgaacgcgtatcacccgtggtctaaagatacaga 2519
DB 2461 TTACCCTTAAAGAAAGAACGATGATCAGCCCAACGCTATCACCGTGGCTAAAGATACGA 2520
QY 2520 caaagaatgccgcgcagtcagcatctaaactcaaaagtaaaaaacgctctaaacggttg 2579
DB 2521 CAAAAATGCCGCGCAGTCAGCATCTTAAACTCAAAAGTAANAACGGCTCAACGGTTG 2580
QY 2580 ctacaaaaaagatggtacggttaccttggccttagccaagatagcgtctgaccattg 2639
DB 2581 CTACCAAAAAAGATGTACGGTTACCTTTGGGCTTAGCCAAGATAGCGGTCTGACCATTG 2640
QY 2640 gcaaaagcaccctaacaacagatggttgaactgttaaagataccaacgacaacaatccaag 2699
DB 2641 GCAAAAGCACCTTAAACACAGATGGCTTGACTGTTAAAGATACCAACGAACAATCCAAG 2700
QY 2700 tcggtgctaagtcattaaattactaattgaatgtagtaagtcacagtcactggcattg 2759
DB 2701 TCGGTGCTAATGGCATTAATTTACTAATGTGAATGGTAGTAATCCAGGTACTGGCATTG 2760
QY 2760 caaataccgctgcattaccagagataaaattggccttgcgttctctgattggtgcagttg 2819
DB 2761 CAAATACCGCTCGCATTTACCAGAGATAAAATTGGCTTTGCTGTTCTGATGGTGCAGTTG 2820
QY 2820 atacaacaaaccttatcttgatcaagacaagctacaagttggcaatgttaaagattacca 2879
DB 2821 ATACAACAACCTTATCTTGATCAAGACAAGCTAGCAAGTTGGCAATGTTAAGATTACCA 2880
QY 2880 acactggcatcaacgagctgggttaaagccatcacaggtgtgtcccaacactccttagca 2939
DB 2881 ACACTGGCATTAACGACGGTGGTAAAGCCATCACAGGGTGTCCCCAACACTGCCTAGCA 2940
QY 2940 ttgccgatcaaaagtagccgaacatagaaactggggcaatacaatcccaagacaagaacaaat 2999
DB 2941 TTGCCGATCAAACTAGCGCAACATAGAACTGGGCAATACAATCCCAAGACAAAACAAAT 3000
QY 3000 ccaacgctgccagcattaatatgatataataacaggtcttaacctaaacaaataataaaca 3059
DB 3001 CCAACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAACCTTAAAAAATAATAACA 3060
QY 3060 acccattgacttgcctcaactatgacattgttgacttgcgaatggcaatggcaatggcacca 3119
DB 3061 ACCCATTTGCTTCCACTTATGACATTTGTTGACTTTGGCAATGGCAATGCCACCA 3120
QY 3120 ccgccagtaaacccatgatcacgcgttaacaaaacccgtaaaagtggtatgatgtgaatg 3179
DB 3121 CCGCCACAGTAACCCATGATACCGCTAACCAAAACAGTAAGTGGTATATGATGCTGAATG 3180
QY 3180 tggatatacaaacattatcttaacagggcaactgatgacataaaaaacttggcgtaaaaa 3239
DB 3181 TGGATGATACAACCATTTCACTTAACAGGCACCTGATGACATAAAAAAATTTGGCGTCAAAA 3240
QY 3240 ccacaaactgaacaaaacgaatgctaaatggttaacagcaactaaactttaaattgaact 3299
DB 3241 CCACCAACTGAACAAACAAAGTGTAAATGGTAAATACAGCACTAACTTTAAATGTTAACT 3300
QY 3300 ctagtgtgaagtgccttctttaaagcgaagacatcgccgaaaaatctaaacacccctag 3359
DB 3301 CTAGTGATGAAGATGCCCTTTGTTTAAACGCCAAGACATCGCGGAAAAATCTAAACACCCCTAG 3360
QY 3360 ccaagaaaaattcaacaccacaaaggaacagcagacacccgcttacaacacctttaccgta 3419
DB 3361 CCAAGGAAATTTCAACACCACAAAGGCACAGACACACCGCCCTTACAAACCTTTTACCGTTA 3420
QY 3420 aaaggtagatgaaataataatgctgtagcgcgaacccatcaccggtgggtcaaaaga 3479
DB 3421 AAAAGGTAGATGAATAATAATGCTGATACCGCCRAACGCCATCACCGTGGGTCAAAAGA 3480
QY 3480 acgcaataatcaagtcaaacacccctaacactcaaaaggtgaaaaacggtctttaaattaaaa 3539
DB 3481 ACCGAAATAATCAAGTCAACACCTTAACACTCAAAAGGTGAACGGTCTTAATATTAATA 3540
QY 3540 ccgacaaaaatggtacggtttaccttltggcatataacaccacaagcgttctttaaagccggca 3599

DB 3541 CCACAAAAATGTAGCGTTACCTTTGGCATTAACCAACGCGGTCTTTAAAGCCGGCA 3600
QY 3600 aaagcaccctaaacgacggtggtgtcttatataaaacccactggttagcgaacaaatcc 3659
DB 3601 AAAGCACCTTAACAGCGGTGGTGTCTATTAANAACCCACTGGTAGGACAAATACC 3660
QY 3660 aagtcggtgctgtagcgtgaaatttggccaaggttaataataatggtgtgtgtaggtgctg 3719
DB 3661 AAGTCGTTGCTGATGCGGTGAAGTTTGCCAAAGTTAATAATAATGTTGTTGTTAGTGCTG 3720
QY 3720 gcaattatggcacaactcgcattaccagagatgaaattggctttacttggactaaatggct 3779
DB 3721 GCATTATGTCACAACCTCGCATTTACCAGAGATGAATTTGGCTTTACTGGGACTAATGGCT 3780
QY 3780 cacttgataaaagcaaacccccccttaagcaaacgagcattaacgaggtggttaaaaaaga 3839
DB 3781 CACTTTGATANAAGCAACCCCACTTAGCAAAAGACGGCATTAACGACGGTGGTANAAGA 3840
QY 3840 ttaccaacattcaatcaggtgagattggcccaaacagccatgatcgtgtgacagggcgca 3899
DB 3841 TTACCAACATTTCAATCAGGTGAGATTGCCCCAAAACAGCCATGCTGTGACAGCGGCA 3900
QY 3900 agatttatgatttaaaaaacccgaacttgaacaaacaaatcagcagtagtactgccaacacgac 3959
DB 3901 AGATTTATGATTTTAAACCCGAACCTTGAAAAACAANAATCAGCAGTACTGCCAAAACAGCAC 3960
QY 3960 aaaaactcattacgaattctcagtagcagatgaacaaagtgataaactttacggttagta 4019
DB 3961 AAAACTCATTTACAGCAATTTCTCAGTAGCAGATGAACAAAGGTAAATAACTTTACGGTTAGTA 4020
QY 4020 accttactcagttatgacacctcaaaagacctctgattgcatcacaccttggcaggtgaaa 4079
DB 4021 ACCCTTACTCCAGTTATGACACCTCAAAAGACCTCTGTATGTCATCACCTTTGCAGTGCAA 4080
QY 4080 acggcattacacacaaggttaataaaaggtgtgtggtggtggcattgacacaaacaaag 4139
DB 4081 ACGGCATTTACCAACCAAGGTAAATAAAGGTGTGTGCGTGTGGGCATTGACCAAAACAAAG 4140
QY 4140 gcttaacccagcctaaagctgacgtgggttaataatggtgcaaggtcattgtcattgaca 4199
DB 4141 GCTTAACCAACGCTTAAGTCACCGTGGGTAAATAATATGCAAAAGGCATTTGTCTATTGACA 4200
QY 4200 gccaaaaatgltcaaaaatcaccatcacaggaactgaagcaacacttagctaaattgttaccatg 4259
DB 4201 GCCAAAAATGTTCAAAATACCATCACAGACTAAGCAACACTCTAGCTAATGTTACCAATG 4260
QY 4260 ataaaggtagcgtacgcaccacagaacagggcaataataatcaaaagacgaagacaaaaacc 4319
DB 4261 ATAAAGTAGCGTAGCCACACACAGACAGGGCAATATAATAAAGACGAAGACAAAACCC 4320
QY 4320 gtgccgcagcattgttgatggtctaagcgcaggtttaaacttgcaggtgcaatgggtgaag 4379
DB 4321 GTGCCGCCACATTTGTGATGTGCTAAGCGCAGGCTTTAACTTGCAGAGGCNAATGGTGAAG 4380
QY 4380 cgggtgacttgtctccacttatgacaccctcaacttgcgcatggcaatggcaccaccg 4439
DB 4381 CGGTTGACTTTGTCTCCACTTATGACACCGCTCAACTTTGCGGATGGCAATGCCACACCG 4440
QY 4440 ctaaggtgaacctatgatgacacaagcaaaacccagtaaaagtggctctatgatcgaatggtg 4499
DB 4441 CTAAGGTGACCTATGATGACACAAGCAAAACCCAGTAAGGTGGTCTATGATGTCAATGGG 4500
QY 4500 atgatacaaacattgaagttaaagataaaaaacttgcgttaaaaaaccacacacattgacca 4559
DB 4501 ATGATACAACCATTAAGTTAAAGATAAAAACCTTGCGGTAAAAAACCCACCATTTGACCA 4560
QY 4560 gtactggcacaggtgtgtaataaaatttcccttaagcaatcaagctactggcgatgcgcttg 4619
DB 4561 GTACTGSCACAGGTGCTTAATAAATTTGCCCTTAAGCAATCAAGCTACTGGGATGCGCTTG 4620
QY 4620 tcaagccagtgatatcgttgcctcatcctaacaacaccttatctgtgcacatccaaactgcca 4679

Db 4621 TCAAGGCCAGTGATATCGTTGCTCATCTAAACACACCTTTATCTGGGACATCCAAACTGCCA 4680
Qy aaggggcaagcgaagcgaacaaactcagcaggtctatgttggtgtgtatggtgcaataaagtca 4739
Db 4681 AAGGGCAAGCCAAAGCAACAACTCAGCAGGCTATGTGGATGTGTGATGGCAATAAGGTCA 4740
Qy 4740 tctatgacagtcaccgataaagaagtaactatcaagccaaataatgatggcacagtttgataaaa 4799
Db 4741 TCTATGACAGTACCGATAAACAGTACTATCAAGCCAAATATGATGGCACAGTTGATAAAA 4800
Qy 4800 ccaagaagttgccaagacaaaactggcgcgaagcccaaaccccgatggtgacattgg 4859
Db 4801 CCAACAGATTGCCAAAGACAACTGGTGCGCCCAAGCCCAAAACCCAGATGCGCACATTGG 4860
Qy 4860 ctcaaatgaatgctcaatcagtcattacaagaagaacaaagtaaaatgacaaataaaagc 4919
Db 4861 CTCAAATGAATGTCAAATCAGTATTAAACAAAGAACAAAGTAAATGATGCCAATAAAGC 4920
Qy 4920 aaggcatcaatgaagacacagcccttctgttaaggacttgaagaagccgctctctgataaca 4979
Db 4921 AAGGCATCAATGAAGACAAACGCCCTTGTAAAGGACTTGAAGAGCGCTTCTGATAACA 4980
Qy 4980 aaacaaaacccgcagtaactgttggtgatttaaatgcccgttgcgaacacacgcgtga 5039
Db 4981 AAACAAAACCCGACGTAACGTGTGGGTGATTTAAATGCCGTTGCCCAAAACCCGCTGA 5040
Qy 5040 ccttgcagggatcacaggcacacagcgtcaaaaactggcgagactttgacatcaaaag 5099
Db 5041 CTTTTCAGGGGATACGGGCACACGGCTAAANAACCTGGCGGAGACTTTGACCATCAAG 5100
Qy 5100 gtggcaaacagacacaaataagctaaocgataataacatcogtgtgtgtagcaggtactg 5159
Db 5101 GTGGCAACACAGACACCAATAAGCTAACCGATAATAACATCGTGTGTAGCAGTACTG 5160
Qy 5160 atggtctactgtcaaaacttgcaaaagacctaaaccaacttaacagcgttaagtgcagtg 5219
Db 5161 ATGGCTTCACTGTCAAACTTGGCCAAAGACCTAACCAATCTTAACAGCGTTAATGCAGGTG 5220
Qy 5220 gcaacaaaattgtatgacaaagcgtgtctttgttagactcaagcgggtcaagccaaagcaa 5279
Db 5221 GCACCAAAATTGATGACAAAGGCGTCTCTTTGTAGACTCAAGCGGTCAAGCCAAAGCAA 5280
Qy 5280 acacccctgtctaatggtgcgaactggctgggtgcaagctcatcagtaagtgg 5339
Db 5281 ACACCCCTGTGCTTAAGTGCCAAATGGGCTGGACCTGGGTGGCAAGGTATCATGTAATGTGG 5340
Qy 5340 gcaaaagcacaagaataccgaacgtcgccaaatgtacaaacagtttaaacgaagtacgcaact 5399
Db 5341 GCAAGGCACAAAGATACCGACGCTGCCAATGTACAACAGTTAAACGAAGTACGCAACT 5400
Qy 5400 tgttggtgcttggttaattgctggttaataagataacgctgacggcaatcaggtataacattg 5459
Db 5401 TGTGGGTCTGTGTTAATGCTGGTAATGAATAACGCTGACGGCAATCAGGTAAACATTTGCCG 5460
Qy 5460 acatcaaaaagaccacaaattcaggttcatctcaaccgcactgtcatcaaaagcaggca 5519
Db 5461 ACATCAAAAAGACCCCAAAATTCAGTTTCATCATCTAACCGCACTGTATCAAAAGCAGGCA 5520
Qy 5520 cggtaactggcggtaaaaggtataaacgataaccgaaaaacttgccactggtggtatataaag 5579
Db 5521 CGGTACTTGGCGGTAAAGGTAAATAACGATACCGAAAAAATTTGCCACTGGTGGTATACAAG 5580
Qy 5580 tggcggtgataaagacggaacactaacgcgatttaagcaatggttgggtcaaaaccc 5639
Db 5581 TGGGCGTGGATTAAGACGGCAACCGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAACCC 5640
Qy 5640 aaaaagatggcagcaaaaaagccctgctgcgaactataacgcgcaggttcagacaaact 5699
Db 5641 AAAAGATGGCAGCAAAAAGCCCTGCTCGCCACTTATAACGGCGCAGGTGACAGCAACT 5700
Qy 5700 attgacaaacaccccgcaaaagccattgacagaaataaatgaacaaaggtatccgctct 5759
Db 5701 ATTGACCAACAAACCCGACGAAGCCATTGACAAATAAATGAACACAGGTATCCGCTTCT 5760

Qy 5760 tccatgtcaacgatggcaatcaagagcctgtgtgtacaaagggcgtaacggcattgaactcaa 5819
Db 5761 TCCATGTCAACGATGGCAATCAAGAGCCTGTGTGTAAGAGGCGTAACGCGATTGTACTCAA 5820
Qy 5820 gtgctcaggaagcaactcagtggtgcataggtttccagggccaagcagatggtggaagcgg 5879
Db 5821 GTGCCCTCAGGCAAGCACTCAGTGGGGATAGGTTTTCAGGCCCAAGCAGATGTTGAAGCGC 5880
Qy 5880 cggttgacctagggcagacaaaacccaaagcaggcaacaaatccatcgccatcggtgtaaacg 5939
Db 5881 CCGTTGCCATAGGCAGACAAACCCAAAGCAGCAACCAATCCATCGCCATCGGTGATAAGC 5940
Qy 5940 cacaagccacggggcgatcaatccatcgccatcggttacaggcaatggttagcagtaagc 5999
Db 5941 CACAAGCCACGGGCGATCAATCCATCGCATCGGTACAGGCAATGTGGTAGCAGGTAGC 6000
Qy 6000 actcgtgctcactcgcgacccaagcactgttaagctgtataacagttacagtggtggtta 6059
Db 6001 ACTCTGGTGGCATCGCGGACCCCAAGCACACTTTAAAGCTGATAACAGTTACAGTGTGGTA 6060
Qy 6060 ataacaaccagtttacccgatgccactcaaacccatgtcttctgggtgggcaataacatca 6119
Db 6061 ATAACAACCAAGTTTACCGATGCCACTCAAAACCGATGCTTTGGTGTGGGCAATAACATCA 6120
Qy 6120 ccgtgacccaagaaactcgtgttccttaggttcaaaactctgcatacagtcgagggcacac 6179
Db 6121 CCGTGACCGGAAGTAACTCGGTTGCCCTTAGGTTCAAACTCTGCCATCAGTGCAGGCAC 6180
Qy 6180 acgagggcacacaaagccaaataatctgacggcgacagcaggtacaacacacacagcaggtg 6239
Db 6181 AGGCAGGCACACAAAGCCCAAAATACTGACGGCACAGCAGGTACAAACACACAGCAGGTG 6240
Qy 6240 caacggtaaggttaaaagccttgcgtgacaaacgagcgttgggtgaggtctccgtgggtg 6299
Db 6241 CAACCGGTACCGGTTAAAGGCTTTGCTCGACAAACGGCGGTTGGTGGCGTCTCCGTGGGTG 6300
Qy 6300 cctcaggtgctgaaacgcgtatccaaaatgtggcagcaggtgaggttcagtcgacacagca 6359
Db 6301 CCTCAGGTGCTGAACGCGGTATCCAAATGTGGCAGCAGGTGAGGTGAGTGCACACAGCA 6360
Qy 6360 ccgatgcggtcaaatggttagcaggtgtacaaagccacccaaagcatgccaacgcaacca 6419
Db 6361 CCGATGCGCGTCAATGGTAGCCAGTTGTACAAAGCCACCCAAAGCATTTGCCAACCCACCA 6420
Qy 6420 atgagcttgacacatcgatccacacaaacgaaataaaggccaaatgcagggaggtttcaatc 6479
Db 6421 ATGAGCTTGACCATCGTATCCACCAAAACGAAATTAAGGCCAATGCCAGGCAATTCATCAG 6480
Qy 6480 cgaatggcgatgggtccatgcgaacagcctacattcctcgccagatccatggttaccgggg 6539
Db 6481 CGATGGCGATGGCGTCCATGCCAAGCCTACATTTCTTGGCAGATCCATGGTTACCGGGG 6540
Qy 6540 gtattgcccacccacaaacgggtcaaggtgcggtggcagtgaggactgtcgaagctgtcgagata 6599
Db 6541 GTATTGGCCACCCCAACCGGTCAAGGTGCGGTGGCAGTGGGACTGTTCGAAGCTGTCCGATA 6600
Qy 6600 atggtcaatgggttatttaaaatcaatggttcagccgaatacccaagggccatgtagggggcg 6659
Db 6601 ATGGTCAATGGGTATTTAAATCAATGGTTACGCCGATACCCCAAGGCCATGTAGGGGGCG 6660
Qy 6660 cagttggtgaggttttcaacttttaagccataaaatcgcaagattttacttataaaatcaat 6719
Db 6661 CAGTTGGTGCAGGTTTTCATTTTAAAGCCATAATTCGCAAGATTTTACTTAAAAATCAAT 6720
Qy 6720 ctacacatagttgtataaaacagcatcagcaatcagtcataattactgagctgagtttttt 6779
Db 6721 CTCACCATAGTTGTATAAAAACAGCATCAGTCATATTTACTGATGCTGATGTTTTT 6780
Qy 6780 tatcaacttaaacacattttaccgtcaagtgatctcttccacatgaccaaactcgccatt 6839
Db 6781 TATCACTTAAACCAATTTTACCCTCAAGTGATTTCTTTTACCATGACCAATCGCCATT 6840


```
QY 6840 gatcataggtaaactattagtaaattttatcaatgtatgtttagatagttgtaaaat 6899
|||||
Db 6841 GATCATAGGTAAACTATTAGTAAATTTTATCAATGTAGTTGTAGATAGTTAAAT 6900
|||||
QY 6900 tctgcaattgacaaaatgaccgatttatcccgaaaattctgattgattgacccgttga 6959
|||||
Db 6901 TGTGCCATTGACCAAAAATGACCGATTATCCGAAAATTTCTGATTATGATCCGTTGA 6960
|||||
QY 6960 cctgcagggtgcac 6972
|||||
Db 6961 CTTGCAGGTCGAC 6973
|||||

RESULT 3
AX079913 AX079913 6973 bp DNA PAT 22-FEB-2001
LOCUS Sequence 1 from Patent WO0107619.
DEFINITION AX079913
ACCESSION AX079913
VERSION AX079913.1 GI:13159436
KEYWORDS Moraxella catarrhalis.
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 6973)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of
moraxella
JOURNAL Patent: WO 0107619-A 1 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
source
location/Qualifiers
1..6973
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 2265 a 1555 c 1533 g 1620 t
ORIGIN

Query Match 99.8%; Score 6961; DB 10; Length 6973;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ccatggatatggcagggtgctgcctgcctgcgtatgatggcgatgacacccatttgccc 60
|||||
Db 1 CCATGGATATGGCAGGTGCTCGCGCTGCCGTATGATGCGGATGACACCCCATTTGCC 60
|||||
QY 61 catatctgacgtttgacatgtgatattgaacatgtgacatgatttaacatgttt 120
|||||
Db 61 CATATCTGTACGATTTGACATGTGATATGATTTAACATGTGACATGATTTAACATGTTT 120
|||||
QY 121 aatactgttgccatcattaccataatttagtaacgcatttagtaacgcatttgtaaaaa 180
|||||
Db 121 AATACTGTGCCATCAATTACCATAATTTAGTAACGCATTTAGTAACGCATTTGTAAAT 180
|||||
QY 181 cattggccctttatgtgatcatatgaatagaatattatgattgtatgtatttgt 240
|||||
Db 181 CATGGCCCTTTATGTGTATCATATGAATAGAATAATTATGATTGTATGTATTAATGT 240
|||||
QY 241 atcagaatgtgtatgtatgatgatgcctacagttgatttgggttaactcctcatg 300
|||||
Db 241 ATCAGAATGTTGATGTATGATGATGATGCTACGAGTTGATTGGGTTAATCACTCATG 300
|||||
QY 301 atttgatatatttgaactaatctatttgacttaaatcaccatattggttaatttagca 360
|||||
Db 301 ATTTGATATATTTTGAACATACTATGTACTTAATCAACCATATGTTTATTAATTTAGCA 360
|||||
QY 361 taatgttagcttttggtaaaatcacatgcgaatattgttctactgttactaccatgct 420
|||||
Db 361 TAATGTGTAGCTTTTGTGAAAAATCACATCGCAATATTTGTTACTGTTACTACCATGCT 420
|||||
QY 421 tgaatgacgtcccaatccaccagattcattcaagtgatgtgtgtgtatcagcaccattta 480
|||||
Db 421 TGAATGACGATCCCAATCACACCATTCATTCATAGTATGATGTTGTATACGACCATTTA 480
|||||
```

```
QY 481 ccctaattatttcaatcaaatgectatgtcagcatgtatcatattttttaaggtaaacccac 540
|||||
Db 481 CCCTAAATATTTCATCAATCAAAATGCTATGTAGCATGTATCATTTTTTTTAAAGTAAACCCAC 540
|||||
QY 541 catgaatcacatcatataaadtcatctttaacaaagccacacaggcacatttatggcagtgcc 600
|||||
Db 541 CATGAATCACATCTATAAAGTCATCTTTTAAACAAAGCCACAGGCACATTTATGGCAGTGGC 600
|||||
QY 601 agagtacgccaatacccacagcac-ggggggggtagctgtgtcacaggggcaagtgtgca 659
|||||
Db 601 AGAGTAGCCCAATCCACAGCAGCGGGGGGGTACTGTGCTACAGGGCAAGTTGGCA 660
|||||
QY 660 gtgtatgactctgagctttgtcccgatattgcgcgcgtcgtctgtcctcgtgatcggtgcaa 719
|||||
Db 661 GTGTATGACTCTGAGCTTTGCGCGTATTGCGCGCTCGCTGTCTGCTGATCGTGGTCAA 720
|||||
QY 720 cgtcagtgagcagcttctatgcctcaaaaaaagataccaaacatatcgcaattgtgaac 779
|||||
Db 721 CGTCACTGGCAGTGTCTTATGCTCAAAAAAAGATCAACACATATCGCAATTGGTGAAC 780
|||||
QY 780 aaacacagccaagcagctcagggcactgccaagcgagcggtgtagtcgagccattgtctattg 839
|||||
Db 781 AARACCAAGCCANAGCGCTCAGGCACCTGCCAAGGCGGAGGTGATCGAGCCATTGCTATTG 840
|||||
QY 840 gtgaaaaatgctaacgcacagggcggtccaagccatcgccatcggttagtagtaataaaaactg 899
|||||
Db 841 GTGAAAAATGCTAACGCACAGCGGCGTCAAGCCATCGCCATCGGTAGTAGTAATAAAACTG 900
|||||
QY 900 tcaatggaagcagtttgataaagataggtaccgatgctacgggtcaagagatccatcccca 959
|||||
Db 901 TCAATGGAAGCAGTTTGGATTAAGTAGGTACCGGATGTCAGGGTCAAGAGTCCATGCCCA 960
|||||
QY 960 tcggtgtgtatgtaaaagctagtggtgatgcctcgattgcatcgatcggtgtagtgacttac 1019
|||||
Db 961 TCGGTGTGTATGTAAGGCTAGTGTGTATGCTCGATGCTCGATTGCCATGCTGATGATGACTTAC 1020
|||||
QY 1020 atttgttgatcagcatgtgtaatactcaacatccgaaggtactctgtattaacgatetta 1079
|||||
Db 1021 ATTTGCTGTATCAGCATGCTTAATCTTAACATCCGAAGGTACTCTGATTAAACGATCTTA 1080
|||||
QY 1080 ttaacggcctatgcatattaaaagaaaatacgaagctcaaaagataaagatgataaaataata 1139
|||||
Db 1081 TTAAACGCCCATGTCAGTATTAAAAGAAATACGAAGCTCAAAAGGATAAATGATGTAAATATA 1140
|||||
QY 1140 gacgcacaacccgaagcgcacagccagctactgcagtgaggagccatgctcatatcacagg 1199
|||||
Db 1141 GACGCACAACCGCANAGCGGACACGCCAGTACTGCAGTGGGAGCCATGTCAATATGCACAGG 1200
|||||
QY 1200 gtcatattttccaacgcctttgttacacgggcaacagctaaaaagtgcctatttccttggcag 1259
|||||
Db 1201 GTCATTTTTTCCAACGCCCTTTGGTACACGGGCAACAGCTAAAAAGTGCCCTATTTCCTTGGCAG 1260
|||||
QY 1260 tgggtcttgccgcacaaagcagagggccaaatctacaatcgctatttggttctgatacaaat 1319
|||||
Db 1261 TGGGTCTTGGCGCCACAGCGAGGGCCAATCTACAATCGCTATTGGTTCTGTATGCACAT 1320
|||||
QY 1320 ctgactcgtttggagcagatagcccttggcgaggtactcgtgctcagctacagagggcagta 1379
|||||
Db 1321 CTAGCTGTTGGAGCCGATAGCCCTTGGTGAGGTACTGTGCTCAGCTACAGGGCAGTA 1380
|||||
QY 1380 ttgoccttagtcaaggttctgttgtcactcagatgataataattcttagaccggccctata 1439
|||||
Db 1381 TTGCCCCTAGGTCAAGGTTCTGTGTCTACTCAGAGTGATAATAATTCAGACCGGCCCTATA 1440
|||||
QY 1440 caccataatcccaggcactagaccccccaagtccaagccaccaataataacgaagcggggtc 1499
|||||
Db 1441 CACCAATATCCAGGCACTAGACCCCCCAAGTTTCAAGCCCAACCAATAAATACGAAGCGGGTC 1500
|||||
QY 1500 cactttcattgtgtagtaactctatcaaaagtaaaatcaatcaatgctcggtgaggtgta 1559
|||||
Db 1501 CACTTCCATTGGTAGTAACCTATCAACGTAANAATCATCAATGTCGGTGCAGGTGTTA 1560
|||||
```


QY 1560 ataaacccgatgcgggtcaatgtggcacagctagaacggttggttaagtggcgttaagagc 1619
|||||
Db 1561 ATAAACCCGATGCGGTCAATGTGGCACAGCTAGAGCGGTGGTGAAGTGGCTTAAGAGC 1620
|||||
QY 1620 gtgaattacttttcagtggtgatataacagctactgacgtataaaataggttttgataata 1679
|||||
Db 1621 GTAGATTTACTTTTCAGGGTGATGATACAGTACTGACGTAAAAATAGGTTTGGATAATA 1680
|||||
QY 1680 cttaactattaaaggttggtgcagagacaaacgcattaaacgataataataatcgtgtgg 1739
|||||
Db 1681 CTTTAACATTAAAGGTGGTGCAGAGACCAACGCATTAAACGATTAATAATATCGGTGTGG 1740
|||||
QY 1740 taaagagcgtgaataatagtgctgaaagttaaacttgctaaactttaacaactctta 1799
|||||
Db 1741 TAAAGAGCGCTGATTAATAGTGGTCTGAAAGTTAAACTTTCCTTAAACCTTTAAACAATCTTA 1800
|||||
QY 1800 ctgaggtgaatacaactacattaaatgcccacaacacagtttaaggttaggttagtga 1859
|||||
Db 1801 CTGAGGTGAATACAACTACATTAAATGCGCACAAACACAGTTAAAGTAGGTAGTAGTA 1860
|||||
QY 1860 gtactacagctgaattattagtgatagtttaacctttaccagcgcacataacagcagtc 1919
|||||
Db 1861 GFACTACAGCTGAATTAATGAGTGTAGTTTAACTTTACCCAGCCCAATACAGCAGTC 1920
|||||
QY 1920 aaagcaagcaaaacccgtctatggcgttaaagggtgaaagttaacttaataatgcagaaa 1979
|||||
Db 1921 AAAGCACAAAGCAAAACCCGTCTATGGCGTTAATGGGGTGAAGTTTACTAATAATGCAGAAA 1980
|||||
QY 1980 caacagcagcaatcggcactactcgtattaccagagataaaattggtttgctcgagatg 2039
|||||
Db 1981 CAACAGCAGCAATCGGCACACTACTCGTATTACCAGAGATAAAATTGGCTTGGCTCGAGATG 2040
|||||
QY 2040 gtgatgttataaaacacagcaccatatttgataaaaaaacaacttaagtgggtagtg 2099
|||||
Db 2041 GTGATGTTGATGAAAAAACAAGCACCATATTTGGATAAAAAACAACCTTAAAGTGGGTAGTG 2100
|||||
QY 2100 ttgcaattaccatagacaaatggcattgtatgcagggttaataaaagatcagtaactctggca 2159
|||||
Db 2101 TTGCAATTACCATAGACAAATGGCATTTGATGCAGGTAAATAAAGATCAGTAATCTTGCCA 2160
|||||
QY 2160 aagtgacagctgaacgtcggttaccatgcgaacagctcaaacgcgcaagcctactt 2219
|||||
Db 2161 AAGGTAGCAGTGCTACAGTGCAGTTCACCATCGACAGCTCAAGCGCCAGCCACTACTT 2220
|||||
QY 2220 taaacgagcgcgtgcacatcagtggtcacacacctactgaaataatcagttgtagtgaagtg 2279
|||||
Db 2221 TAAAGCAGCGCGTGGCATCAGTGTACACCTACTGAAATATACGTTGATGTGCTAAGAGTG 2280
|||||
QY 2280 gcaatgttacgcgcccaacttaacatattggctgaaacccacagcttaacagtgatg 2339
|||||
Db 2281 GCAATGTTACCGGCCCAACTTAACATTTGGCGTGAACACCCAGCTTAACAGTGATG 2340
|||||
QY 2340 gcactagtataaatttagtgttaagggtagtggtgcgaacaaatagcttagttaccgcg 2399
|||||
Db 2341 GCACGTAGTAAATTTAGTGTAAAGGTAGTGTGACGAACATAGCTTAGTTACCGCG 2400
|||||
QY 2400 aacatttggcgaagctatctataatgaagtcgaatgaacgcgtgcagctgtctacaagct 2459
|||||
Db 2401 AACATTTGGCAAGCTATCTAAATGAAGTCAATGAACCGCTGACAGTGTCTCAAAAGCT 2460
|||||
QY 2460 ttaccgttaagaagaagacagatgatgcgcccaacgctatcacgltggcttaagaatcacga 2519
|||||
Db 2461 TTACCGTTAAAGAAGAAGACGATGATGACCCCAACGCTATCACCGTGGCTTAAGATACGA 2520
|||||
QY 2520 caaaaaatgcgcgcagtcagcatcttaaaactcaaaaggttaaaaaacggtctcaacggttg 2579
|||||
Db 2521 CAAAAATGCGCGCGCAGTCAGCATCTTAAACTCAAAAGGTAAAAACGGTCTTAACGGTTG 2580
|||||
QY 2580 ctacccaaaaagatgggtacggttaccttttggttaacgaagatacgggtctacacattg 2639
|||||
Db 2581 CTACCAAAAAAGATGGGTACGGTTTACCTTTGGCGTTACCCCAAGATAGCGGTCTGACCAATTG 2640
|||||
QY 2640 gcaaaagcacccctaaacaaagtggtgactgttaagaatcaccaacgaacaaatccaaag 2699
|||||

Db 2641 GCAAAAGCACCCCTAAACACGATGGCTTGACTGTTAAAGATACCAACGAACAAATCCAAG 2700
|||||
QY 2700 tcsgtctaagtcgcatataattacttaattgaatgtagtaatccaggtactcgtggtcattg 2759
|||||
Db 2701 TCGGTGCTAATGCGATTAATTTACTTAATGTGAATGGTAGTATCCAGGTACTGGCATTG 2760
|||||
QY 2760 caaatccgctcgcattaccagagataaaattggctttgctgttctctgagtggcagttg 2819
|||||
Db 2761 CAATACCGCTCGCATTTACCAGAGATAAAATTTGGCTTTGCTGTGATGGTGCAGTTG 2820
|||||
QY 2820 atacaaacaaaccttatcttgatcaagagaagctacaagttggcaatgttaagatcacca 2879
|||||
Db 2821 ATACAAACAACCCCTTACTTTGATCAAGACAAGCTACAAAGTTGGCAATGTTAAAGTTACCA 2880
|||||
QY 2880 acactggcattaacgcaggttggtaagcccatcacaggtgtgcccaacactgcctagca 2939
|||||
Db 2881 ACCTGGCATTTACGAGAGTGGTAAGCCCATCACAGGCTGTCCCCAACACTGCCTAGCA 2940
|||||
QY 2940 ttgcgcatcaaaagtagccgcaacatagaactgggcaatatacaatcccaagacaagaacaaat 2999
|||||
Db 2941 TTGCGGATCAAAAGTAGCCGAACATAGAACTGGGCAATACAAATCCAAGACAAAGACAAAT 3000
|||||
QY 3000 ccaacgctgcagcatttaataatgatatataaacacagcctttaacctataaaaaataaaca 3059
|||||
Db 3001 CCAACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAACCTTAAAAATAATAAACA 3060
|||||
QY 3060 acccattgactttgtctccacttatgacattgttgactttggcaatggcaatggccacca 3119
|||||
Db 3061 ACCCATTTGACTTTGTCTCCACTTATGACATTTGTGACTTTGGCAATGGCAATGCCACCA 3120
|||||
QY 3120 ccgcaacagtaaccatgatacccgtaacaaaacagtaaaagtgtgtatgtatgagtgaatg 3179
|||||
Db 3121 CCGCACAGTAAACCCATGATACCGCTTAACAAACAGTAAAGTGGTATATGTGTGTAATG 3180
|||||
QY 3180 tggatgatacaaacattcatctaacagcactgatgacaataaaaaacttggcgtcaaaa 3239
|||||
Db 3181 TGGATGATACAAACCATTCATCTAACAGCCTGATGACAAATAAAAACTTGGCTGCAAAA 3240
|||||
QY 3240 ccacaaactgaacaaacaaagtgctaatggtaataacagcaactaaacttttaattgaact 3299
|||||
Db 3241 CCACCAAACTGAACAAACAAGTGCTTAATGGTAAATACAGCAACTAACTTTAATGTTAACT 3300
|||||
QY 3300 ctagtgaagaatgccttgttaacgcaaaagcatcgccgaataatctaaacacccctag 3359
|||||
Db 3301 CTAGTGTAGTGAAGTGCCTTGTTAACGCCAAAGACATCGCCGAAAACTTAAACACCTTAG 3360
|||||
QY 3360 ccaaggaattcacaccacaaagggcacagacacacgcgcctacaaacccctttaccggtta 3419
|||||
Db 3361 CCAAGGAAATTCACACACCACCAAGGCACAGCAGACACCGCCCTACAAACCTTTACCGTTA 3420
|||||
QY 3420 aaaaagtgatgaataataatgtctgatgcgcgaacgcctacacccgtgggtcaaaaga 3479
|||||
Db 3421 AAAAGGTAGTGAATAATTAATATGCTGATGACGCCCAAGCCCATCACCGTGGGTCAANAAGA 3480
|||||
QY 3480 acqcaaaaatacaagtaacaccccttaacactcaagggtgaaggggttcttaataataaaaa 3539
|||||
Db 3481 ACGCAATAATACTAAGTCAACACCCCTTAACACTCAAAAGGTGAAACCGTCTTAATATTAAAA 3540
|||||
QY 3540 ccgcaaaaaatgttcaggttacctttggcatttaacacacagcgcgtctctaaagccgca 3599
|||||
Db 3541 CCGACAAAATGGTACGGTTACCTTTGGCATTTAACACCAACAGCGCTCTTAANGCGCGCA 3600
|||||
QY 3600 aaagcaccttaaacgagcgtggctgtcttataaaaaacccactggtgagcgcaacaaatcc 3659
|||||
Db 3601 AAAGCACCTTAACGACGCTGGCTTGTCTATTAAAAACCCCACTGGTAGCAACAAATCC 3660
|||||
QY 3660 aagtcggtgctgatggcgtgaaggttgcgaagtttaataataatgggtgtgtaggtgctg 3719
|||||
Db 3661 AAGTCGCTGCTGATGGCGTGAAGTTTGGCAAGGTTAAATAATATGCTGTGTAGGTGCTG 3720
|||||
QY 3720 gcaatgaatggcacaactcgcattaccagagatgaatggctttactgggactaaaggct 3779
|||||

Db 3721 GCATTGATGGCACAACATCGCATTTACAGAGATGAATTTGGCTTTACTGGGACTAATGGCT 3780
 QY 3780 cacttgataaaagcaaacccacccttaagcaaaagacgcatcattaaagcaggtgggtataaaga 3839
 Db 3781 CACTTGATAAAGCAAAACCCACCCTAAGCAAAACACGCGATTAACGCGAGGTGGTAAAAAGA 3840
 QY 3840 ttaccacattcaatcaggtgaagatggcccaaaagacgatgatgctgtgacaggggca 3899
 Db 3841 TTACCACATTCATCAGGTGAATTTGCCCAAAACACGCCATGATGTTGACAGCGGCA 3900
 QY 3900 agatttatgattlaaaaaacgaaactgaaaaaacaatacagcagctactgccaacaaacagcac 3959
 Db 3901 AGATTTATGATTTAAAAACGCACTTGAAAAACAATAATCAGCAGTACTGCCAAAACAGCAC 3960
 QY 3960 aaactcattacagcaattctcagtagcagatgaacaaggaataaactttacggttagta 4019
 Db 3961 AAAACTCATTFACAGGAATCTCAGTAGCAGATGAACAAGGTAATAACTTTACGGTTAGTA 4020
 QY 4020 acccttactccagttatgacacctcaaaagacctctgatgcatcaacctttgagggtaaa 4079
 Db 4021 ACCCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCATCACCTTTGCAGGTGAA 4080
 QY 4080 acgacattaccaccaaagtaataaagggtgtgtgcgtgggcatcattgaccaaaccacaaag 4139
 Db 4081 ACCGCAATACCAACAAGGTAATAAAGGTGTGTGGCTGTGGCCATTTGACCAACCAAAAG 4140
 QY 4140 gcttaaccacgcctaagctgacgctgggtaataataatggcaagcattgtcatigaca 4199
 Db 4141 GCTTAACCAACGCTAAGCTGACCGTGGGTAATAATAATGCAAGGCAATGTCTATTGACA 4200
 QY 4200 gccaaatggtcaaaataccatcacaggactaaggcaacaccttagctaaatggttaccaatg 4259
 Db 4201 GCCAAATGTGCAAAATFACCATCACAGGACTAAGCAACACTCTAGCTAATGTATACCAATG 4260
 QY 4260 ataaagtagctacgcaccacagacagggcaataataatcaaaagcagaaacaaaccc 4319
 Db 4261 ATAAAGTAGCTACGCCACCAAGACAGGCGCAATATAATCAAGACAGACAAAACCC 4320
 QY 4320 gtgcccccagcattgttgatgtctgaagcagcagctttaaacttgcaaggcaatggtaag 4379
 Db 4321 GTGCCCGCAGCATTTGTGATGTGCTAAGCGCAGGCTTTAACTTTGCAAGGCAATGGTGAAG 4380
 QY 4380 cgttgacttgtctccacttatgacaccgtcaactttgcggtgcaatggcaccacgg 4439
 Db 4381 CGGTTGACTTTGTCTCCACTTATGACACCGCTCAACTTTGCCGATGCGCAATGCCACCACCG 4440
 QY 4440 ctaaggtgacctatgatgacacaaagcaaaaccagtaaaagtgtctatgatgtcaatgtgg 4499
 Db 4441 CTAAAGGTGACCTATGATGACACAAAGCAAAACCAAGTAAAGTGGTCTATGATGTCAATGTGG 4500
 QY 4500 atgatacaaccattgaagttaaagataaaaaacttggcgtaaaaaaccacacattgacca 4559
 Db 4501 ATGATACAACCATTTGAAGTTAAAGATAAAAAAATTTGGCGTAAAAACCAACCCACCATTTGACCA 4560
 QY 4560 gtactggcacaggtgtataataatttggccctaaggcaatacaagctactggcgatggccttg 4619
 Db 4561 GTACTGGCACAGGTGTATTAATAATTTGCCCTAAGCAATCAAGCTACTGGCGATGGCGGTTG 4620
 QY 4620 tcaagggcagtgatagttgttgcactcaaacaccttatctggcgacatccaaactggcca 4679
 Db 4621 TCAAGGCCAGTGATATCGTTGCTCATCTAAACACCTTATCTGGCGACATCCAAACTGCCCA 4680
 QY 4680 aaggggcaagcgaacaaactcagcaggtctatgtggatgctgaaggcaataaaggtca 4739
 Db 4681 AAGGGCAAGCCAAAGCAACACTCAGCAGGCTATGTGGGATGCTGATGGCAATAAAGGTCA 4740
 QY 4740 tctatgacagtacgataacaaagtactatcaagccaaaaatgatggcaaggtgataaaa 4799
 Db 4741 TCTATGACAGTACCGATAACAAGTACTATCAAGCCAAAAATGATGGCACAGTTGTGATAAAA 4800
 QY 4800 ccaaaaagttgcaaaagcaaaactggtcgcccaagcccaaaacccagatggcacattgg 4859
 Db 4801 CCAAAAGATTGGCAAAAGCAAACTTGTTCGCCCAAGCCCAACCCCAAGATGGGCACATTTGG 4860

QY 4860 ctcaaatgaatgtcaaatcagtcatttaacaagaacaagttaaattgatgtccaaataaaagc 4919
 Db 4861 CTCAAAATGAATGTCAAATCAGTCATTAAACAAGAACAAAGTAATGATGCCAATAAAAAAGC 4920
 QY 4920 aagcgcataatgaagacaacgccttctgttaagagacttgaaaaagcgcgtcttgataaca 4979
 Db 4921 AAGGCATCAATGAAGACAAACGCGCTTTGTTAAAGACTTTGAAAAAGCGCTTCTGTATAAAC 4980
 QY 4980 aaaccaaaaacgcgcagtaactgtgggtgatttaaatgcggttgcccaaacacccgctga 5039
 Db 4981 AAACCAAAAACGCGCAGTAACCTGTGGTGATTTAAATGCCGTTCGCCCAACACCGCTGA 5040
 QY 5040 cctttgcaggggtatagcaggcacaacggcgtaaaaacttggcgagactttgaccatcaaaag 5099
 Db 5041 CCTTTGCAGGGGATACAGGCACAAACGCTAAAAAAGCTGGCGAGACTTTGACCATCAAG 5100
 QY 5100 gtgggcaaacagacacacaaataagcctaaccgataataacatcggtgtggtagcaggtactg 5159
 Db 5101 GTGGGCAAAACAGACACCAATAAGCTAACCAGATAATAACATCGGTGTGGTACGAGGTACTG 5160
 QY 5160 atggcttcactgtcaaaacttggcaaaagacctaaccatcttaacagcgttaatgcaggtg 5219
 Db 5161 ATGGCTTCACTGTCAAACTTTGCCAAAGACCTTAACCAATCTTTAACAGCGTTTAATGCAGGTG 5220
 QY 5220 gcacaaaaatttgatgacaaaaggcggtgctctttttagactcaagcggtcaagcccaaaacaa 5279
 Db 5221 GCACCAAAATTTGATGACAAAGGCGTGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAA 5280
 QY 5280 aacccctgtctgaatggcaatgggctggacctgggtggcaagggtcatcagtaattggtg 5339
 Db 5281 ACACCCCTGTGCTTAAGTGGCAATGGGCTGGACCTGGGTGGCAAGGTCAATGTAATGTGG 5340
 QY 5340 gcaaaagcacaaaagataccgacgctgccaatgtacacacagttaaacgaaagacgcacact 5399
 Db 5341 GCAAAAGCACAAAAGATACCGACGCTGCCAATGTACAAACAGTTAAACAGATACCGCAACT 5400
 QY 5400 tgttgggtcttggtaattgtctggtaattgataacgtcagcgcaatcaggtaaacattgccc 5459
 Db 5401 TGTGTGGTCTTGTGTAATGCTGGTAATGATAAGCTGACGGCAATCAGGTAAACATTGCCG 5460
 QY 5460 acatcaaaaaagaccccaaatcagggttcacatcataaccgcgactgtcatcaaaagcagga 5519
 Db 5461 ACATCAAAAAGACCCCAAAATTCAGGTTTCATCATCTAAACCCGACTGTTCATCAAGACGGCA 5520
 QY 5520 cgttacttggcgtgtaaaaggttaataacgatcccaaaaacttggcacttgggtatacaag 5579
 Db 5521 CGGTACTTGGCGGTAAAGGTAAATAACGATACCGAAAAAATTTGCCACTTGGTGGTATACAAG 5580
 QY 5580 tggcggtggataaagcagcgcaacgcgtacacggtatttaagaatgtttgggtcaaaaaccc 5639
 Db 5581 TGGCGGTGGATAAAGACGGCAACGCTAACGGCGATTTAAGCAATGTTTGGGTCAAAAACCC 5640
 QY 5640 aaaaaagtgacgcaaaaaagccctgctgccacttataacgcgcaggtcagaccacact 5699
 Db 5641 AAAAAGATGCGACAAAAAGCCCTGCTCGCCACTTATAACGGCGGAGGTTCAGACCAACT 5700
 QY 5700 atttgacaaacaccccgacagagccattgacagaaataatgaacaggtatccgctctct 5759
 Db 5701 ATTTGACCAACACCCCGCAGAACCCATTGACAGAAATAATGAACAAGGTATCCGCTTCT 5760
 QY 5760 tccatgtcaacgatggcgaatcaagagcctgtgggtacagggcggttaacggcattgactcaa 5819
 Db 5761 TCCATGTCAACAGATGGCAATCAAGAGCCTGTGGTACAAAGGGCGTAAACGGCATTGACTCAA 5820
 QY 5820 gtgcctcagcagcagcactcagtggtgtagtttccaggccaaagggcagatggtgaagccg 5879
 Db 5821 GTGCCTCAGGCAAGCACTCAGTGGCGATAGTGTTCAGGGCCAAAGCGAGATGGTGAAGCCG 5880
 QY 5880 ccgtgtccatagcgacagacaaaccccaagcaggaacacatccatcgccatcggtgataacg 5939
 Db 5881 CCGTTGCCATAGGCAGACAAAACCCCAAGCAGCAACCACTCCATCCCATCGGTGATTAACG 5940

RESULT 4			
QY	5940	cacaagccacgggagatgaatccatccatcgccatcggtacaggcaaatgtgtgtagcaggtaagc	5999
DB	5941	CACAAGCCACGGGAGATGAATCCATCCATCGCATCGGTACAGGCAATGTGTAGCAGGTAAGC	6000
QY	6000	actctgtgaccatcggcgacccagcactgttaaggctgataaacagttacagttggtgta	6059
DB	6001	ACTCTGTGGTCCATCGCGACCCAGCACTGTTAAGGCTGATACAGTTACAGTCTGGGTA	6060
QY	6060	ataacaacagtttacgatgccactcaaaccgatgtcttttgggtggtggtggtggtggtg	6119
DB	6061	ATAACAACAGTTTACGATGCCACTCAACCGATGTCTTTGGTGTGGCAATAACATCA	6120
QY	6120	ccgtgaccgaagaactcgttgccttagtttcaaaactctgcatcagtcagtcagtcagtc	6179
DB	6121	CCGTGACCGAAGTAACTCGCTTGGCTTAACTCTCAAACTCTGCCATCAGTGCAGGCACAC	6180
QY	6180	acgcagggcacacaagcccaaaaaatctgacggcacagcaggtacacaccacacacagcgtg	6239
DB	6181	ACGCAGGCACACAGCCAAAAAATCTGACGGCACAGCAGGTACAAACACACACAGCAGGTG	6240
QY	6240	caaccggttacggttaaaagcttctgtgacaaaacggcggttgggtgcggtctccgtgggtg	6299
DB	6241	CAACCGGTACGGTTAAAGGCTTTGCTGACAAACGGCGGTTGGTGCGGTCTCCGTGGGTG	6300
QY	6300	cctcaggtgctgaacgcgcgtatccaaaaatgtggcagcaggtgaggtcagtcgcccaccgca	6359
DB	6301	CCTCAGGTGCTGAACGCCGTATCCAAATGTGGCAGCAGGTGAGGTCAGTGCCACCAACA	6360
QY	6360	ccgatgcggtcaatggtgagcagttgtacaaagccaccacaaagcaatgtccacgcacaca	6419
DB	6361	CCGATGCGGTCAATGCTAGCCAGTTGTACAAAGCCACCACCAAGCAATTGCCAACGCAACA	6420
QY	6420	atgagcttgacacatcgtatccacaaaacgaaaataagcccaatgcaggtatctcatcag	6479
DB	6421	ATGAGCTTGACCATCGTATCCACCAACGAAATTAAGCCCAATGCAGGGATTCNTCAG	6480
QY	6480	cgatggcagtgctgcacagcctacatctctggcagatccatgggtaccgggg	6539
DB	6481	CGATGCGATGGCTCCATGCCACAGCCTACATCTCTGGCAGATCCATGTTACCGGGG	6540
QY	6540	gtattgccaccacaaacggtcaaggtgcggtggcagtgaggactgtcgaagctgtcggata	6599
DB	6541	GTATTGCCACCCACAAACGGTCAAGTGGGTGGCAGTGGGACTGTGCAAGCTGTCCGGATA	6600
QY	6600	atgggtcaatgggtatttaaaatcaatgggttcagccgataccacaaagccatgtaggggcgg	6659
DB	6601	ATGGTCAATGGGTATTTAAATCAATGGTTCAGCCGATACCCCAAGCCATGTAGGGCGG	6660
QY	6660	cadttggtgcaggttttccactttaagccataaatgcgaagattttactttaaaatcaat	6719
DB	6661	CAGTTGCTGCAGGTTTTCACCTTTAAGCCATAAATGCGAAGATTTTACTTAAAAATCAAT	6720
QY	6720	ctcaccatagtataaaacagcatcagcatcagtcataattactgactgctgattttt	6779
DB	6721	CTCACCATAGTTGTATAAACACGATCAGATCAGTCATATTTACTGATGCTGATGTTTT	6780
QY	6780	tatcacttaaacattttaccgtcaagtgtattctctttcaacatgaccacaaatcgccatt	6839
DB	6781	TATCACTTAAACCATTTTACCCTCAAGTGATTCTCTTCCACCATGACCAAAATCGCCATT	6840
QY	6840	gatacaggtaaacattatgtagtaaaattttatcaatgtagttgttagatatgttaaat	6899
DB	6841	GATCATAGGTAACCTATTACAGTAATTTTATCAATGTAGTTGTTAGATAGGTTAAAT	6900
QY	6900	tgtgcaattgacaaaataagcagatttatcccgaataattctgattatgatccgttga	6959
DB	6901	TGTGCCATTGACCAAAAAATGACCGATTATCCCGAAAAATTTCTGATTATGATCCGTTGA	6960
QY	6960	ccgtcaggtcgac	6972
DB	6961	CCTGCAGGTGCAC	6973

RESULT	4				
LOCUS	AX079924	6259 bp	DNA	PAT	22-FEB-2001
DEFINITION	Sequence 12 from Patent WO0107619.				
ACCESSION	AX079924				
VERSION	AX079924.1	GI:13159442			
KEYWORDS	Moraxella catarrhalis.				
SOURCE	Moraxella catarrhalis.				
ORGANISM	Moraxella catarrhalis				
REFERENCE	1 (bases 1 to 6259)				
AUTHORS	Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.				
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella				
JOURNAL	Patent: WO 0107619-A 12 01-FEB-2001;				
FEATURES	CONNAUGHT LABORATORIES LIMITED (CA)				
source	Location/Qualifiers				
	i. .6259				
	/organism="Moraxella catarrhalis"				
	/db_xref="taxon:480"				
BASE COUNT	2067 a	1414 c	1393 g	1385 t	
ORIGIN					
	Query Match 89.8%; Score 6258; DB 10; Length 6259;				
	Best Local Similarity 100.0%; Pred. No. 0;				
	Matches 6258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	708	tgatcggtcaacgctcagtcgagtcgtcttctgcataaaaaaagataccaaacatctcg	767		
DB	2	TGATCGGTCAACCGCTCAGTGGCGAGTCTTATGCTCAAAAAAAGATACCAACATATCG	61		
QY	768	caattggtgaacaaaacagccaagacgctcagcactgcgaagggcagcgggtgacgag	827		
DB	62	CAATTGTTGAACAAACCCAGCAGCGCTCAGGCACTGCCAAGGGGAGCGGTGATCGAG	121		
QY	828	ccatgctattggtgaaaaatgctaaccacagggcggtcaagccatcgcatcggtagta	887		
DB	122	CCATTGCTATTGTTGTAATAATGCTAACGCACAGGGCGGTCAAGCCATCGCCATCGGTAGTA	181		
QY	888	gtaataaaactgcaatgggaagcagtttggataagataggtaccgctgctacgggtcaag	947		
DB	182	GTAATAAAACTGTCAATGGGAAGCAGTTTGGATAAGATAGGTACCGATGCTACGGGTCAAG	241		
QY	948	agtcacatgccatcgggtggtgtagtataaaggctagtggtgtagctcgtatgccatcgta	1007		
DB	242	AGTCATCGCCATCGGTGGTGATGATAAAGGCTAGTGGTGTATGCCATCGATGCCATCGGTA	301		
QY	1008	gtgatgacttaccatttctgtgatcagcatggttaatcctaaacatccgaaggttactctga	1067		
DB	302	GTGATGACTTACATTTGCTTGTGATCAGCATGGTAACTCTTAACATCCGAAGGTACTCTGA	361		
QY	1068	ttaacgatcttattaaacggccatcagttattataaagaataacgaagctcaagataatg	1127		
DB	362	TTAACGATCTTATTAAACGGCCATGCAGTATTAAAAAGAAATACGAAGCTCAAGGATTAATG	421		
QY	1128	atgtaaaatataagacgcaac	1187		
DB	422	ATGTAAAAATATAGACGACACACCCGCAAGCGGACACGCCAGTACTGTCAGTGGAGCCCATGT	481		
QY	1188	catatgcacaggtctatttttccaaacgcttttgggtacacggggcaacagctaaaagtgcct	1247		
DB	482	CATATGCACAGGGTCATTTTTCACGCTTTGGTGTACACGGGCAACAGCTAAAAAGTGCCT	541		
QY	1248	attccttggcagtggtgttgcgcacacagcgggccaatctacacgtctattgtt	1307		
DB	542	ATTCTTGGCAGTGGGTCTTTGCCGCCACAGCGGAGGGGCCAATCTACAAATCGCTATTGGTT	601		
QY	1308	ctgatgcacacatctagctcgtttgggagcgatagcccttgggtgtaggtactcgtctcagc	1367		
DB	602	CTGATGCAACATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGCAGGTACTCTGCTGCACG	661		

QY 1368 tacaggcgagatgtcccttaggtcaagggttctgtgtcactcagagtgataataattcta 1427
 Db 662 TACAGGCGAGATGTGCCCTTAGGTCAAGGTTCCTGTGTCTACTCAGAGTGATAATAATTCTA 721
 QY 1428 gacggcctatacacaataaccaggcgactagaccacccaggtttcaagccaccacaataa 1487
 Db 722 GACCGGCTATACACCAATACCCAGGCACTAGACCCCAAGTTTCAAGCCACCAATAATA 781
 QY 1488 cgaaggcggtccactttccattttagtagtaactctatcaaacgtataaatcatcaatgtcg 1547
 Db 782 CGAAGCGGGTCCACTTTTCCATTGGTAGTAACCTATCAAAACGTAAATCATCAATGTGC 841
 QY 1548 gtacagggttaataaaccagatgcgggtcaatgtgcacagctagaaagcggttgtaagt 1607
 Db 842 GTGACGGTGTAAATAAACCGATGCGGTCAATGTGCACAGCTAGAAAGCGGTGTGAAGT 901
 QY 1608 gggctaaggagcgtagaattacttttcagggtgatataacagctactgacgtataaaatag 1667
 Db 902 GGCTAAGGAGCGTAGAATTTACTTTTCAGGGTGATGATAACACTACTGACGTAANAATAG 961
 QY 1668 gtttgataataactttaactattaaagggtgtgcagagaccaacgcatttaaccgataata 1727
 Db 962 GTTTGGATAATACTTTAACTATTAAAGGTGTCAGAGACCAACGCCATTAAACCGATAATA 1021
 QY 1728 atatcggtgtggttaaaagaggtgataataatgtggtctgaagttaaaacttgctaaaaact 1787
 Db 1022 ATATCGGTGTGTTAAAGAGGCTGATAAATAGTGTCTGAAAGTTAAACTTTGCTAAAACTT 1081
 QY 1788 taacaacttactgaggtgaatatacaactacattaaatgcccacaacacaggttaaggtag 1847
 Db 1082 TAACAATCTTACTGAGGTGAATACAACTACATTAATGCCCACAAACCACAGTTAAGGTAG 1141
 QY 1848 gtatgtagtagtactacagctgaattattgtagtgatagtttaaccctttaccagcccca 1907
 Db 1142 GTAGTAGTAGTACTACAGCTGAATATTAGTGTAGTACTTTAACTTTACCAGCCCA 1201
 QY 1908 atacaggcagtaaaagcacaagcaaacccgtctatggcgtaagtgggtgaagttaacta 1967
 Db 1202 ATACAGCAGCTCAAGACACAAAGCAAAACCCTCTATGCGGTTAATGGGGTGAAGTTTACTA 1261
 QY 1968 ataatgcagaacaacagcagcaatcgccactactcgtattaccagagataaaattggct 2027
 Db 1262 ATAATGCAGAAACACAGCAGCAATCGGCACACTACTCGTATTACAGAGATAAATTTGGCT 1321
 QY 2028 ttgctcgagatgtgtatgttgataaaaaacaagcacccatatttgggataaaaaacaactta 2087
 Db 1322 TTGCTCGAGATGCTGATGTTGATGAAAAACAAGCACCATATTTTGGATAAAAAACAACCTTA 1381
 QY 2088 aagtgggtagttgtaattaccatagacaatggcattgtatgcaggttaataaaaaagatca 2147
 Db 1382 AAGTGGGTAGTGTGCAATTTACCATAGACAAATGGCATTTGATGAGSTAAATAAAGATCA 1441
 QY 2148 gtaacttcgaaggtagcagtgtcaacgatcggtttaccatcgcaacagctcaaaagcg 2207
 Db 1442 GTAATCTTGCAAAAGGTAGCAGTGTCAACGATGCGGTTACCATTGACACGCTCAAGCCG 1501
 QY 2208 ccaagcctaactttaacgcgagcggtggcctcagtgctcacacctactgaaatatcagttg 2267
 Db 1502 CCAAGCCTACTTTAAACCGCAGGCGCTGGCATCAGTGTCAACCTTACTTGAAATATCAGTTG 1561
 QY 2268 atgctaagagtggaatgtttaccgcccccaacttacaacatlggctgtgaaaacacccagc 2327
 Db 1562 ATGCTAAGAGTGGCAATTTTACCGCCCCCACTTACAACATTTGGCGTGAACACCCAGC 1621
 QY 2328 ttaacagtgtgcaactagtgataaaatttagtgttaagggtaggtgtacgaacaatagct 2387
 Db 1622 TTACAGTGTATGACACTAGTGTATAAATTTAGTGTTAAGGGTAGTGTACCAACATAGCT 1681
 QY 2388 tagttaccgcccgaacttttggcagcttatctaataatgaagtcaatcgacggctgacaggt 2447
 Db 1682 TAGTTACCGCCGAACAATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGGCTGACAGTG 1741
 QY 2448 ctctaaaaagcitttaccggttaaaagaagaacgatgatgacgccaacgctatcacogtgg 2507

Db 1742 CTCTACAAAGCTTTACCGTTAAAGAAGAAGCAGATGATGCGCAACGCTATCACCGTGG 1801
 QY 2508 ctaaaagatagacaaaaatgcccggcgagtcagcagctcttaaaactcaaaaggttaaaacg 2567
 Db 1802 CTAAAGATAGACAAAANAATGCCGCGCAGTCAGCATCTTAAAACTCAAAAGGTAAANAAG 1861
 QY 2568 gtctaacgggtgtctaccacaaaaagatgtacgggttaacctttgggcttagcacaagatagcg 2627
 Db 1862 GTCTAACGGTTGCTTACCAAAAAAGATGTCAGGTACCGTTACCTTTGGGCTTTAGCCAAGTAGCG 1921
 QY 2628 gtctgaccatttgcaaaagcacccttaaaacacgatggcttgactgtttaaagataaccaacg 2687
 Db 1922 GTCTGACCATTTGGCAAAAGCACCCTAAACAACAGATGCTTGACTGTTTAAAGATACCAACG 1981
 QY 2688 acaaaatccaagtctggtgctcaatgtgcattaaatttactaatgtgaatggttagtaataccag 2747
 Db 1982 AACAAATCCAAAGTCGTTGCTAATGGCAATTAATTTACTAATGTGAATGGTAGTAATCCAG 2041
 QY 2748 gtactggcattgcaaaataccgctcgcattaccagagataaaaattggcttggcttggcttctg 2807
 Db 2042 GTACTGGCATTTGCAAAATACCGCTCGCATTACCAGAGATAAAATTTGGCTTTGCTGTTCTG 2101
 QY 2808 atggtgcagttgatatacaaaacaccccttatcttgcatacgaagcaacgctcaaatgtggcaatg 2867
 Db 2102 ATGGTGCAGTTGTATACAAAACAAACCTTATCTTGATCAAGACAAGCTCAAAAGTTGGCAATG 2161
 QY 2868 ttaagattaccaacacactggcattaaacgcaggtggtgaagccatcacagggtgctgcccaa 2927
 Db 2162 TTAAAGATTACCAACACTTGGCATTAACCGAGGTGGTAAAGCCNTCCAGGGCTGTGCCCAA 2221
 QY 2928 cactgcctagcattgcccgaatcaaaagtagccgcaacatagaaactgggcaatcaaatccaaag 2987
 Db 2222 CACTGCCTAGCATTTGCCGATCAAAAGTAGCCGAACATAGAACTGGGCAATACAATCCAAG 2281
 QY 2988 acaaaagcaaatccaaagctgcagcgaattaatatataataacaggctttaaacctaa 3047
 Db 2282 ACAAAAGCAAAATCCAGCTGCCAGCATTAATGATATATTAAATACAGGCTTTAAACCTAA 2341
 QY 3048 aaaaataaacaaccccatgtacttctccacttatgacattgttgacttttgcccaatg 3107
 Db 2342 AAAATATACAAACCCCATTTGACTTTTCTCCACTTATGACATTTGTTGACTTTGCCAATG 2401
 QY 3108 gcaatgccacccgcgcacagtaaacccatgataccgctaaacaaacccagtaaaagtgtat 3167
 Db 2402 GCAATGCCACCCGCCACAGTAACCCCATGATACCGCTAACAAAAACAGTAAGTGGTAT 2461
 QY 3168 atgatgtgaatgttgatgatatacaaacattcatctaaacaggcactgatgacaataaaaaac 3227
 Db 2462 ATGATGTGAATGTGGATGATACAAACCATTCATCTAAACAGGCACGTGATGACAAATAAANAAC 2521
 QY 3228 ttggcgtcaaaacccacaaactgaacaaacgaagtgttaattggttaatacacagcaactaact 3287
 Db 2522 TTGGCGTCAAAACCCACCAACTGAACAAACAAAGTGTGTAATGTGAATACAGCACTACT 2581
 QY 3288 ttaatgttaactctagtgatgaagatgcccttggtaaagcgaagacagatgccgaaatc 3347
 Db 2582 TTAATGTTAACTCTAGTGTAGTGAAGATGCCCTTTGTTAAACGCCAAAAGACATCGCCGAAAATC 2641
 QY 3348 taaaaccccttagcgaaggaataatcaaccccaaaagccacagacagacccgcctatacaaa 3407
 Db 2642 TAAACACCCCTAGCCAAAGAAATTCACACCAACCAAGGCACAGCAGACCCGCCCTTACAAA 2701
 QY 3408 cctttaccgttaaaaaggttagtgaataataatgtctgtagcgccaacgccatcacg 3467
 Db 2702 CCTTTACCGTTAAANAAGGTAGTGAANAATAATATGCTGATGACGCCNACGCCATCACCG 2761
 QY 3468 tgggtcaaaagacgcgaataataatcaagtcacacccctaaactcaaaaggtgaaacggctc 3527
 Db 2762 TGGGTCAAAAAGACGCAAAATAATCAAGTCAACACCCCTAAACACTCAAAAGGTGAAAACGGTC 2821
 QY 3528 ttaatatataaacacgcgaacaaaatgttagcggttaccctttggcatttaaacacacagcggtc 3587

QY 1022 ttgcttgatcagcgtgtaataccttaacatcccgaaaggtactctctgattaacgactttatt 1081
 Db 481 TTGCTTGTATGACATGTAATCTTAAACATCCGAAAGGTACTCTGTGATTAACGATCTTATT 540
 QY 1082 aacggccatcagattataaagaataacgaagctcaagataatgatgtaaaatataga 1141
 Db 541 AACGGCCATGCGATATTAAAGAATAACGAGCTCAAGGATATGATGTAAATATAGA 600
 QY 1142 cgcacaacgcaagcgagacacgcagctactgagtcagtcggagccatgtcatatgcacaggg 1201
 Db 601 CGCACACCGCAAGCGGACACCGCCAGTACTGTCAGTGGGAGCCATGTCTATATGACAGGGT 660
 QY 1202 catttttccacgcctttggttacacgggcaacagctaaagtgccattctctggcagtg 1261
 Db 661 CATTTTCCAAACGGCTTTGGTACACGGGCAACAGCTAAAGTGCGCTATTCTCTGGCAGTG 720
 QY 1262 ggtcttgccgcacagcgaggggccaactctacaatcgctattggttctctgatgcaacatct 1321
 Db 721 GGTCTTGGCCGACACGAGCGGAGGCCAATCTACAAATCGCTATTGTTCTGATGCAACATCT 780
 QY 1322 agctgttggagcgatagcccttgggtcaggttactcgtcagctacagggcagttatt 1381
 Db 781 AGCTCGTTGGGAGCGATAGCCCTTTGGTGCAGGTACTCGTGTCACTACAGGGCAGTATT 840
 QY 1382 gccctagatcaaggtctctgttgcactcagagtgataataattctagaccggcctataca 1441
 Db 841 GCCCTTAGGTCAAGGTTCTGTGTCACTCAGAGTGATAATTTCTAGACCGGCTATACA 900
 QY 1442 ccaataccccggcactagaccccaagtttcaagccacccaataatacgaagcgcggtcca 1501
 Db 901 CCNAATACCCAGGCACTAGACCCCAAGTTTCAAGCCACCAATAATACGAGCGGGTCCA 960
 QY 1502 ctttccattggttagtaactctatacaacgtaaaatcatactcgttcggtcaggtgttaatt 1561
 Db 961 CTTTCCATTTGGTGTAGTAACCTATCAACCTAATAATCAATCAATCGGTGCGAGGTGTTAAT 1020
 QY 1562 aaacacgaagcggtcaattgtgcacagcttagaagcggtgggtaagtcggttaagcggt 1621
 Db 1021 AAACCGATGCGGTCAATGTGTCACAGCTAGAACGGGTGGTCAAGTGGGCTAAGGAGCGT 1080
 QY 1622 agaattactttcaggtgtagataacagctactcagctaaagtaaaataggtttgataatact 1681
 Db 1081 AGAATTTACTTTTCAGGGTGATGATAACAGTACTGACGTAAATAATAGTTTGGTAAATACT 1140
 QY 1682 ttaactattaaaggtggtgcagagacccaacgcatcattaccgataataatcgcgttggtta 1741
 Db 1141 TTAACCTATTAAAGGTGTCAGAGACCAACGATTAACCGATAAATAATATCGGTGTGTA 1200
 QY 1742 aaagagctgataatagtggtctgaaagttaaacttgctaaacitttaacaaacttact 1801
 Db 1201 AAAGAGGCTGATAATATGCTGTGAAGTTAAACTTGGCTTAAACCTTTTAAACAACTTTACT 1260
 QY 1802 gaggtgaatacactacattaaatgcacaacacagtttaaggtaggttagtagtagtagt 1861
 Db 1261 GAGGTGAATACAACTACATTAATGTCACACACACACAGTTAAGGTAGGTAGTAGTAGT 1320
 QY 1862 actcagctgaattatttgatgtagtttaacocctttaccagcccaataacagcgagtcasa 1921
 Db 1321 ACTACAGCTGAATTAATTAGTGTAGTTTAAACCTTTTACCAGCCCAATAACAGGCAGTCAA 1380
 QY 1922 agcacaagcaaaacccgtctatggtctgaaagtggtggaagtttacttaataatgcagaaca 1981
 Db 1381 AGCACAAGCAAAAACCGTCATGCGGTTAATGGGGTGAAGTTTACTAATTAATGCAGAAACA 1440
 QY 1982 acagcagcaatcgccactactctattaccagagataaaattggctttgctcgagatggt 2041
 Db 1441 ACAGCAGCAATCGCACTACTCTGATTTACAGAGATAAATAATGGCTTTGCTCGAGATGTT 1500
 QY 2042 gatgttgataaacaagcaccatatttgataaaaaaacacttaagtggttagttgtt 2101
 Db 1501 GATGTTGATGAAAAAAGCACCATATTGATGATAAAAAACCTTAAGGTGGGTAGTGT 1560
 QY 2102 gcaattaccatagacaatgscattgatgcagggttaataaaaaagatcagtaattcttgccaaa 2161

Db 1561 GCATTTACCATACACAATGGCATTGATGAGGTAAATAAAGATCAGTAATCTTGCACAA 1620
 QY 2162 ggtagcagtgctaaacagatgcggtttaccatcgaaacagctcaagccgcaagcctacttta 2221
 Db 1621 GGTAGCAGTGCTTAACGATGCGTTTACCATCGAACAGCTCAAGCCGCCAAGCCTACTTTA 1680
 QY 2222 aacgcaggcgctggcatcagtgccacacctactgaaatatcagttgagtcgaagagtgcc 2281
 Db 1681 AACGAGCGCTGGCATCAGTGTACACCTTACTGAAATATACGTTGATGATGATGATGATG 1740
 QY 2282 aatgttaccgcccacttacaacattggtgtaaaacccacccagcttcaacagtgatggc 2341
 Db 1741 AATGTTACC GCCCAACTTACAACATTGGCGTGAACACCCACGAGCTTAACAGTGATGGC 1800
 QY 2342 actagtgataaaatttagtgtaaggttagtggtacgaacaatagcttagttaccgcccga 2401
 Db 1801 ACTAGTGATTAATTTAGTGTAAAGGTAGTGGTACGNACAAATAGCTTAGTTACCGCGGA 1860
 QY 2402 cattggcaagctatctaaatgaagtcaatcgaaacgggtgacagtgctctacaaaagcttt 2461
 Db 1861 CATTTGGCAAGCTATCTTAAATGAAGTCAATCGAACGGCTGACAGTGTCTTACAAAGCTTT 1920
 QY 2462 accgttaaaagaagaacgatgatgacgccaacgctatcacccgtggctaaagatcacaca 2521
 Db 1921 ACCGTTAAAGAAGAAGACGATGATGACGCCARCGCTATCACCGTGGCTAANGATACGACA 1980
 QY 2522 aaaaatcccgcgagtcagcatcttaaaactcaaaaggtaaaaacgctctaaacggttgc 2581
 Db 1981 AAAAATGCCGGCGAGTCAGCATCTTAAACTCAAAAGGTAAACACGGCTTAACCGGTTGCT 2040
 QY 2582 accaaaaagatggtacggtttacccttggccttagccaagatagcgtctgaccattggc 2641
 Db 2041 ACCAAAAAGATGTTAGCGTTTACCTTTGGGCTTAGCCAAAGATAGCGGTCTGACCACTTTGGC 2100
 QY 2642 aaagacccttaaaacagatggtcttgactgttaagatacccaacgcaaaaatccaaagtc 2701
 Db 2101 AAAAGACCCCTTAACACAGTGGCTTGACTGTAAAGATACCAACGCAAAATCCAAGTC 2160
 QY 2702 ggtgctaattggtcaattacttaattgtaattggtgtaattccaggttactggtgcatgca 2761
 Db 2161 GGTGCTAATGGCATTAATTTACTTAATGTGAATGTGAATGTGAATGTGAATGTGAATGTGA 2220
 QY 2762 aatccctcagctaccagagataaaattggttggcttctggttctgagtggtgagtgat 2821
 Db 2221 AATACCGCTCGCATTTACCAGAGATAAATTTGCTTTGCTGTTCTGATGTTGATGATGAT 2280
 QY 2822 acaaaacaccccttatcttctgatacagaacagctcaagttggcaatgttaagattacacaac 2881
 Db 2281 ACAAAACAACTTATCTTGTATCAAGACAAGCTACAAGTTGGCAATGTTAAGATTACCAAC 2340
 QY 2882 actggcataaacgaggtggttaaagccatcacaggtctcccccacactgccttagcatt 2941
 Db 2341 ACTGGCATTTAAGCAGGTGGTAAAGCCATCACAGGGCTGTCCCCNACACTGCCTAGCATTT 2400
 QY 2942 gccgatcaagtagcgcgcaacatagaactgggcaatacaataccaaagacaaagacaaatcc 3001
 Db 2401 GCCGATCAAAAGTAGCGGCAACATAGAACTGGGCAATAACAATCCAAGACAAAGACAAATCC 2460
 QY 3002 aacgctgcagcatttaattgatatataataacaggtcttaaccttaaaaaataataacaac 3061
 Db 2461 AACGCTGCCAGCATTAAATGATATATTAATACAGGCTTTAACCTTAAATAAATAAATACAAAC 2520
 QY 3062 cccattgacttggctccacttatgacattggtgactttggcaatggcaatggccaccacc 3121
 Db 2521 CCCATTGACTTTGCTCCACTTATGACATTTGACTTTGCCAATGGCAATGCCACCACC 2580
 QY 3122 gccacagtaaacatgataccgcctaacaacacagtaaaagtggttatatgattgattggtg 3181
 Db 2581 GCCACAGTAACCCATGATACCGCTAACAAAACAGTAAGAGTGGTATATGATGATGATGATG 2640
 QY 3182 gatgatacaaccattcatcacaagcactgatgacaataaaaaaacttggtcgtcaaaacc 3241

Db 2641 GATGATACAAACCAATTCATCTAACAGGCACTGATGACAATAAATAAACTTGGCGTCAAAACC 2700
QY 3242 accaaactgaacaaaacagtgctaataatggttaatacacagcaactaacctttaatgttaactct 3301
Db 2701 ACCAACTGAACAAAACAAGTGCTAATGGTAAATACAGCAACTAACTTTAATGTTAACTCT 2760
QY 3302 agtgaatgaagatgccttgttaacgcgaagacatcgcgcgaataatcgaacacccctagcc 3361
Db 2761 AGTGATGAAGATCCCTGTGTTACGGCAAGACATCGCCGAAATCTAAACACCCTAGCC 2820
QY 3362 aaggaattcacaccaccacaaagccagcagacacccgcctcacaacccctttaccgttaa 3421
Db 2821 AAGGAATTCACACCAACCAAGCACAGACACCCGCCCTACAAACCTTTACCGTTAA 2880
QY 3422 aagtgatgaaaaataatgctgatgcgcgaacgcacatcacccgtgggtcgaagaac 3481
Db 2881 AAGGTAGATCAAAATAATATGCTGATGACGCCAACGCCATCACCGTGGTCAAAAGAAC 2940
QY 3482 gcaaatataacagtcacaccccttaacactcaaaaggtgaaacacggtcttataatataaac 3541
Db 2941 GCAAAATAATCAAGTCAACACCCCTTAACACTCAAAAGGTGAAACCGGTCTTAATATTA 3000
QY 3542 gacaaaaatggtacggttacctttggcattacacacacagcggtcttaaacgcgcga 3601
Db 3001 GACAAAATGTTAGGTTACCTTTGGCATTTAACCAACAGCGGTCTTTAAGCCGGCA 3060
QY 3602 agcacctaaacagcggtgctgtctattaaaaacccactgtagcgaacaaatccaa 3661
Db 3061 AGCACCTTAAACGACGCTGCTGTTCTATTAAAAACCCCACTGGTAGCGAACAATCCAA 3120
QY 3662 gtcggtgctgagtgagttgcaagtttcaaggttaataataatggttctgtagtgcgtg 3721
Db 3121 GTCGGTGCTGATGCGGTGAAGTTTGCAAGGTTAAATAATATGCTGTGTAGTGTGCTGC 3180
QY 3722 attgatggcacaactcgattaccagatgaattggctttactggactaatgactca 3781
Db 3181 ATTGATGGCAACACTCGATTACAGAGTGAATTTGGCTTTTACTGGGACTAATGGCTCA 3240
QY 3782 cttgataaagcaaacccacacctaagcaagacggtattaacgcaggtggtgtaaaagatt 3841
Db 3241 CTTGATAAAGCAAAACCCCACTTAAGCAAGACGGCATTAACGACGCTGTTAAAGATT 3300
QY 3842 accaaattcaatcaggtgagatgccccaaaacagccatgatgctgtgacagcggcgaag 3301
Db 3301 ACCAACTTCAATCAGTGAGATGGCCCAAAACAGGCATGATGCTGTGACAGCGGCAAG 3360
QY 3902 attatgatttaaaacccgaactgaaacaaataacagcagtaactgccaaaacagcaaa 3961
Db 3361 ATTTATGATTTAAACCCGAACTTGAAACAAATAACAGCATGCTGCTGCTGCTGCTGCTG 3420
QY 3962 aactattacagaattctcagtagcagatgaacaaggttaataactttacggttagtaac 4021
Db 3421 AACTATTACACCAATTCACATGACAGATGAACAAGGTAATAACTTTACGGTTAGTAAC 3480
QY 4022 ccttactccagttatgacacctcaagacactcgtatgctatcacccttgcaggtgaaaac 4081
Db 3481 CCTTACTCCAGTTATGACACCTCAAGACCTCTGATGCTATCATCACTTTGACGCTGAAAC 3540
QY 4082 ggcattaccacaagtaataaaggtggtgctggtggtggtggtggtggtggtggtggtggt 4141
Db 3541 GGCATTACCAACCAAGGTAAATAAAGGTGTGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 3600
QY 4142 ttaaccacgcctgaactgacgtgggttaaataatggcgaagcattgtcattgacagc 4201
Db 3601 TTAACCAACGCCTAAGCTGACCGTGGGTAAATAATGGCAAGGCAATGTGTCATTGACAG 3660
QY 4202 caaaatggtcaaaatcaccatcacagactaagcaacactctagctaatgttacaatgat 4261
Db 3661 CAAAATGTTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTTACCAATGAT 3720
QY 4262 aaaggttagcgtacgcacacagacagggcaataataatcaaacgaagacaaacccgt 4321
Db 3721 AAAGGTAGCGTACGCACCAAGAACAGGGCAATATAATCAAGACGAAAGAACAAACCCGT 3780

QY 4322 gcgcgcagcattgtgatgtgtaagcgcaggttttaacttgcaagcaatggtggaagcg 4381
Db 3781 GCCCCAGCATTTGTGATGTGCTAAGCGCAGGCTTTAACTTGAAGCAAGCAAGCG 3840
QY 4382 gttgaacttgcctccacttatgacacccgtcaactttgcgcagatggcaatgcaccacccgt 4441
Db 3841 GTTGACTTTGCTCTCCACTTATGACACCGCTCACTTTGCCGATGCAATGCAACCGCT 3900
QY 4442 aaggtgacctatgatgacacaagaacacacagtaaaagtgtcttatgtatgtaattgtgat 4501
Db 3901 AAGGTGACCTTATGATGACACAGCAAAACCGTAAAGTGGTCTATGATGTAATGTGGAT 3960
QY 4502 gatacaaccttaagttaaagataaaaaacttggcgtaaaaaacccacacattgaccagt 4561
Db 3961 GATACAACCAATTTGAAGTTAAAGATAAAAACTTGGCGTAAAAACCCACCATTTGACCAGT 4020
QY 4562 actggcacaggtgctaataatttgcctcaagaacactcaactactggcgatggccttgc 4621
Db 4021 ACTGGCACAGGTCTAATAAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGGCGTTGTC 4080
QY 4622 aaggcagtgatgctgtctcatcctaaacacacttctgcgcacatccaaactgcga 4681
Db 4081 AAGGCCAGTGTATGCTGTCTCATCTTAACACCTTTATCTGGCGACATCCAAACTGCCAAA 4140
QY 4682 ggggcaagcccaagcaaacactcagcaggttatgtggtatgctgatggcaataaagtcac 4741
Db 4141 GGGCAAGCCCAAGCAACAACTCAGCAGGCTATGTGGATGCTGATGCAATAAGGTGTCATC 4200
QY 4742 tatgacagtcacagataaacagtgactatcaagcgaataatgatggcacagttgataaac 4801
Db 4201 TATGACAGTACCGATACAAAGTACTATCAAGCCAAAAATGATGGCACAGTTGATAAAC 4260
QY 4802 aaagaagtgcacaaagcaaacactggtgcgcgaagcccaaacccagatggcacattgct 4861
Db 4261 AAAGAAGTTCGCAAGCAACAACTGCTGCCAAGCCCAACCCGATGGCACATTGGCT 4320
QY 4862 caaatgaattcaaatcagtcattacaagaacaaatgaatgataatgataatgataaagca 4921
Db 4321 CAAATGAATGCAAACTCAGTCATTAAACAAAGCAACGTAATAATGATGCAATAAAGCA 4380
QY 4922 ggcataatgaagaacacgccttgttaaggacttgaagaagcgccttctctgataacaa 4981
Db 4381 GGCATCAATGAAGACACAGCGCTTTGTTAAAGGACTTGAAAGAGCGCTTCTGTATAACAA 4440
QY 4982 accaaaacccgcagtaactggtggtgatttaaatccggttgcccaaacacgcgtgacc 5041
Db 4441 ACCAAAACCCGCGAGTAACCTGCTGGGTGATTTAATGCCGTGGCCCAACACCGCTGACC 4500
QY 5042 ttgcaagggtatcacaggcaacacgcgttaaaaaactggcgagactttgaccatcaaggt 5101
Db 4501 TTTGCAAGGGATACAGGCACACAGCGCTAAAAAATGGCGAGACTTTGACCATCAAGGT 4560
QY 5102 gggcaaacagacaccaataagctaaaccgataataaacatcggtgtgtagcaggttactgat 5161
Db 4561 GGGCAACACAGACACCAATAAGCTAAACCGATAATAACATCGGTGTGTGTGAGCAGGTACTGAT 4620
QY 5162 ggttcactgtcaaaactgcgaagacacacacacacacacacacacacacacacacacacac 5221
Db 4621 GGCTTCTACTGTAAACTTGCCAAAGACTAACCACTAACCACTTAACAGCGTTAATGCAAGGTGGC 4680
QY 5222 accaaaattgatgacaagcggtgtctttttagactcaagcggtcaagcgaacacacacacacac 5281
Db 4681 ACCAAAATTGATGACAAGCGGTGCTCTTTTGTAGACTCAAGCGGTCAAGCCCAAGCAAC 4740
QY 5282 acccctgtgctaagtgcaatggcgctggacccctggcgcaaggttcacgttaattgtgggc 5341
Db 4741 ACCCCTGTGTAAGTGCATAATGGCTGGAGCTGGAGTGGCGAAGGTGATCAGTAATGTGGGC 4800
QY 5342 aaaggcacaaagatcacgcgctgccaatgtacacagittaaacgaagtcagcaacttg 5401
Db 4801 AAAGGCACAAAGATACCGAGCTGCCAATGTACACAGTTAAACAGGTAACGAAAGTACGCAACTTG 4860

Qy	5402	ttgggtcttgtaatactggttaataagcgtgacggcaatcaggtaaacattgcgcac	5461		
Db	4861	TTGGGTCTTGTAATGCTGTAATGATAACGCTACGGCAATAGGTAAACATTTGCCGAC	4920		
Qy	5462	atcaaaaaagcccaaatctcaggttcatctcaacgcgactgtcatcaaaagcaggcacg	5521		
Db	4921	ATCAAAAAGACCANAATTCAGGTTTCATCTCAACGCACTGTCTCATCAAGACAGGCACG	4980		
Qy	5522	gtacttggcgttaaggttaataacgatacagaaatactgccactggttgatatcaagt	5581		
Db	4981	GTACTTTGGCGTTAAAGGTAATTAAGCATACCGAAAAACTTGCACCTGGTGTATACAAGTG	5040		
Qy	5582	ggcgtgataaagcgcacgcctaaacggcgtatttaagcaatgtttgggtcaaaacccaa	5641		
Db	5041	GGCGTGGATAAAGACGGCAACGCTAACGGCCATTTAAGCAATGTTGGGTCAAAACCCAA	5100		
Qy	5642	aaagatggcagcaaaaaagccctgctgccacttataacgcgcgaggttcagaccactat	5701		
Db	5101	AAAGATGCGCAGCAAAAAGCCCTGCTCGCCACTTATAACGCGCAGGTTCAGACCAACTAT	5160		
Qy	5702	ttgacaacaaccccgacgaagccattgacagaataaataagcaaggtatccgctcttc	5761		
Db	5161	TTGACCAACAACCCCGCAAGCCATTGACAGATAAATGAACAAAGGTATCCGCTTCTTC	5220		
Qy	5762	catgtcaacgatggcaatcaagagcctgtgtacagggcgtaacggcattgactcaagt	5821		
Db	5221	CATGTCAAGCATGGCAATCAAGAGCCTGTGTGTACAAAGGCGTAAACGGCATTTCAAGT	5280		
Qy	5822	gcctcaggcaacatcactgctggcgttagtttccagggccaagggcagatggtgaagcgc	5881		
Db	5281	GCCTCAGGCAAGCACTAGTGGCGTAGTGTTCAGGCAAGGCAGATGGTGAAGCCGCC	5340		
Qy	5882	gttcctatgacagcaaaacccaagcagcaaccaatccatcgccatcggtgataacgca	5941		
Db	5341	GTTCGCATAGGCAGACAACCCCAAGCAGCAACCAATCCATCGCCATCGGTGATAACGCA	5400		
Qy	5942	caagccacggcgatcaatccatcgccatcggtgacaggaatgtggtgacaggttaagcac	6001		
Db	5401	CAAGCCACGGCGATCAATCCATCGCCATCGGTACAGGCAATGTGTAGCAGGTAAAGCAC	5460		
Qy	6002	tctgtgcatcggcagcaccagcactgttaagctgataacagttacagttggtgtaat	6061		
Db	5461	TCGTGTGCATCGGCAACCCNAGCACTGTTAAGCTGATTAACAGTTACAGTTGGTGAAT	5520		
Qy	6062	aacaacagttaccgatgccactcaaacccgatgttcttgggtggtgggcaataacatcac	6121		
Db	5521	AACAACCAAGTTTACCAGTCCACTCAACCCGATGTCTTTGGTGTGGGCAATAACATCAC	5580		
Qy	6122	gtacacgaagaactcgttgccttagttcaaacctctgccatcagtgacggcacacac	6181		
Db	5581	GTACCCGAAGTAACTCGGTTGCTTTAGGTTCAAACTCTGCCATCATAGTCAGGCACACAC	5640		
Qy	6182	gcagggcacacaagcccaaaaaatctgcgcgcagcagcaggttacacccaccacagcgtgca	6241		
Db	5641	GCAGGCACACAGCCAAAATCTGACGCACAGCAGGTACAAACCAACCAACAGCAGGTGA	5700		
Qy	6242	accggtacggttaaaagccttgtgacaaaacggcgttgggtggtctccgtgggtgcc	6301		
Db	5701	ACCGGTACGGTTTAAAGCCTTTGCTGCAAAACCGCGGTGTGGTGCCTCTCCGGTGGGTGCC	5760		
Qy	6302	tcaggtgtcgaacgcgcatcccaaatgtgcagcagcaggtgaggtcagtgccaccagcac	6361		
Db	5761	TCAGGTGCTGAACCCGCTATCCAAAATGTGGCAGCAGGTGAGGTCAAGTCCACCAACACC	5820		
Qy	6362	gatcgggtcaatggtagccagttgtacaagcccaacccaagcattgcccaacgcaacaaat	6421		
Db	5821	GATCGGTCAATGTTAGTAGCTGTGTACAAAAGCCACCAAGCAATTGCCAAACGCAACCAAT	5880		
Qy	6422	gagcttaccatcgtatccaccacaaacgaataaagcccaatcagggatttcacagcg	6481		
Db	5881	GAGCTTGACCATCGTATCCCAACCAACCAAAATTAAGGCCAATCCAGGGATTTTCATCAGCG	5940		
Qy	6482	atggcagtggttcattgcacaaagcctacattccttggcagatccattggttaccgggggt	6541		

QY	1127	gatgtaaatatatagacgcacaaacccgcaagcgagacacgcccagtgctgagtgaggccatg	118
DB	421	GATGTAAATATATAGACGACAACCGAAGCGAGCAGCGCAGTACTGTGAGTGGGAGCCATG	480
QY	1187	tcatgcacagggctcatctttccaacgcctttgtgtacacgggcaacagctcaaaagtgc	1246
DB	481	TCATATGCACAGGGTCATTTTTCCAAAGCCTTTGGTACACGGGCACAGCTTAAAGTGC	540
QY	1247	tattccttggcagtgggcttgcgcacacagccgagggccaaatctcaaatcgctattgg	1306
DB	541	TATTCTTTGGCAGTGGGTCTTCGCGCCACAGCCGAGGGCCAACTCTACAATCGCTATTGGT	600
QY	1307	tctgatgcaacatctagctctgttgggagcgatagcccttggtcaggtactctgctcag	1366
DB	601	TCTGATGCAACATCTAGCTGTGGGAGCGATAGCCCTTGGTCAGGTACTCTGTCGTCAG	660
QY	1367	ctacagggcagttaccccttaggtcaaggtctctgttgcactcagagtataaattct	1426
DB	661	CTACAGGCGAGTATTCGCCCTAGGTCTAAGGTCTTGTTGTCACAGAGTGATAAATTCT	720
QY	1427	agaccggcctatacacaaatacccaggcacttagaccccaagtttcaagccaccaataat	1486
DB	721	AGACCGGCTTATACACAAATACCCAGGCACCTAGACCCCNAGTTTCAAGCCACCAATAAT	780
QY	1487	acgaaggcggtccactttccattggtagtaactctataaacgtataaatcatcaatg	1546
DB	781	ACGAAGCGGGTCCACTTTCCATTGGTAGTAAGTCTATCAACGTAATAATCATCATGTC	840
QY	1547	ggtcaggtgttaataaaaccgatcggtccaatgtggcacagctagaaagcgggtgtaag	1606
DB	841	GGTCAGCGGTGTAATAAAACCGATGGGTTCATGTGGCACAGCTAGAACGGGTGGTGAAG	900
QY	1607	tgggtcaaggagcgtagaattacttttcagggtgtagataacacgactacgacgtaaaaata	1666
DB	901	TGGGCTAAGGAGCGTAGAATTACTTTTTCAGGGTGATGATACAGTACTGAGTAAAAATA	960
QY	1667	ggtttggataatactttaactataaagtggtgcagagaccaacgcattaaccgataat	1726
DB	961	GGTTTGGATATACTTTTAATATTAAAGGTGGTCAGAGACCAACGCATTTAACGGATAAT	1020
QY	1727	aatacggctgtgataaagagcgtgataatagtgctgaaagttaaaacttgctaaaact	1786
DB	1021	AATATCGGTGTGTAAAGAGCGTGATAATAGTGTGTCGAAAGTTAACTTTGCTAAAACT	1080
QY	1787	ttaacaatcttactgaggtgaatacaactacatlaaatgccacaaccacagtttaaggt	1846
DB	1081	TTAAACAATCTTACTGAGGTGAATACAACACTACATTAATAGCCACACACAGTTAAGSTA	1140
QY	1847	ggtagtagtagtactacagcgtgaattattgtagttagtattaaactttaccagccc	1906
DB	1141	GGTAGTAGTAGTAGTACTACAGCTGAATTTATTGAGTGTAGTTTAACTTTTACCCAGCCC	1200
QY	1907	aatacggcagtcacaagcaagcaaaacggctcgtatggcgttaattgggtgaaagttaact	1966
DB	1201	AATACAGCGAGTCAAGACACAGCAAAACCGTCTATGGCGTTAATGGGGTGAAGTTTACT	1260
QY	1967	aataatgcagaacaacacagcagcaatcggcactactcgtattaccagagataaaattggc	2026
DB	1261	AATAATCGAANAACACAGCAGCAATCGGCCTACTCTGTATTACCGAGATAAATTGGC	1320
QY	2027	tttgctcgatgggtgatgttgatgaaacaagcaccataatttggataaaaaacaact	2086
DB	1321	TTTGCTCGAGATGGTGATGTTGATGAAAAAACAAGCACCACTATTTGGATAAAAACAACCT	1380
QY	2087	aaatgggttagtgttgcattaccatagacaatggcatgtgatgaggtataaaaaagatc	2146
DB	1381	AAAGTGGGTAGTGTGCAATTACCATAGACAATGCCATTGTATGCGAGTAATAAAAAAGATC	1440
QY	2147	agtaatcttgccaaaagtagcagtcacagatcggttaccatcgcaacagctcaagacc	2206
DB	1441	AGTAATCTTCCCAANAGTAGCAGTGTACAGATCGGTTTCCATTCGACACAGCTCAAGACC	1500
QY	2207	gccaaagctactttaaacgcaggcgtggcatcgagtgtcaacactactgaaatcaagt	2266

db	1501	 GCCAAGCCTACTTTTAAACGAGCGCTGGCATCAGTGTCCACACCCTACTCTGAATATCATGTT	1560
Qy	2267	gatctaagagtggcaatgttacccgcccaacttaaacattggcgtgaaaaaccacgag	2326
Db	1561	GATGCTAAGAGTGCATGTTACCGGCCCACTTACAACTTTGGCGTGAACACCAACCGAG	1620
Qy	2327	cttaacagtgatggcactagtcataaatttagtttaaagggtagtggtacgaacaatagc	2386
Db	1621	CTTAACAGTGATGCACTACTGATATAAATTAGTGTTAAGGGTAGTGGTACGAACAATAGC	1680
Qy	2387	ttagttaccgccgaacatttggcgaactatctaaatgaagtcoactgaacggctgacagt	2446
Db	1681	TTAGTTTACCGCGCAACATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGCGTCACAGT	1740
Qy	2447	gctctaaaaagctttaccgtttaaagaagaagacgatgatgacgccaaacgctatcacgtg	2506
Db	1741	GCTCTCAAAAGCTTTACCGTTTAAAGAAGAACGATGATGAGCCCAACGCTATCACCGTG	1800
Qy	2507	gctaaagatacgacaataaagtcggcgagtcagcatcttaaaactcaaaaggtaaaaaac	2566
Db	1801	GCTTAAGATACGACAAAAATGCCGGCAGTCAGCATCTTAAACTCAAAAGGTAAAAAC	1860
Qy	2567	ggcttaacgggttgctaccaaaaaagatggtaacggttaccttgggcttagcacaagatagc	2626
Db	1861	GGTCTAACGGTTGCTACCAAAAAAGATGGTAGGGTTACCTTTTGGCTTAGCCAAGATAGC	1920
Qy	2627	ggctgaccattggcaaaaacacccctaaacaacagatggctgactgtttaagataccaac	2686
Db	1921	GGTCTGACCATTTGGCAAAAGACCCCTAAACAACGATGGCTTGACCTTTAAAGATACCAAC	1980
Qy	2687	gaacaaatccaaagtcggtgctaagtgcatataaatttaactaatgtgaatgtagtaatcca	2746
Db	1981	GAACAATCCAAGTCGGTGTATFGGCATTAAATTTACTATGTCGAATGGTAGTAATCCA	2040
Qy	2747	ggctactggcattgcaaatcacgctcgcattaccagagataaaattggctttgctgggtct	2806
Db	2041	GGTACTGGCATTTGCMAATACCGCTCGCATTTACCAGAGATAAAATTTGGCTTTGCTGGTCT	2100
Qy	2807	gatggtcagttgatcaaaaacaaaccttatcttgatcaagacaagctacaagttggaacat	2866
Db	2101	GATGGTCAGTTTGATCAAAACAACCTTTATCTTGATCAAGACAAGCTACAAAGTTGGCAAT	2160
Qy	2867	gttaagattaccaaacctggcattaacgcaggtgtgtaaaagccatcacaggcgtgcccca	2926
Db	2161	GTTAAGATTACCAACTGSCATTACGCAGGTGGTAAAGCCATCACAGGGCTGTCCCCA	2220
Qy	2927	acactgctagcttggccgatcaaaagtagccgcaacatagaactgggcaatcacatccaa	2986
Db	2221	ACACTGGCTAGCATTTGCCGATCAAAAGTAGCCGCAACATAGAACTGGGCAATACAATCCA	2280
Qy	2987	gacaagacaaaatccaaagctgccagcataatgatataataaatacagggtttaacctta	3046
Db	2281	GACAAAGCAAAATCCAACGCTGCCGATTAATGATATATTAATATACAGGCTTTAACTTA	2340
Qy	3047	aaaaataatacaaccccatgtgcttgcctcaacttatgacattgttgactttgccaat	3106
Db	2341	AAAAATATACAAACCCATTGACTTTGTCTCCACTATGACATTGTTGACTTTGGCCAT	2400
Qy	3107	ggcaatgccaccacccgccaagtaaccccatgatcacctgataccgctaaacaaaccagtaagtggta	3166
Db	2401	GGCAATGCCACCACCGCCAGTAACCCATGATACCGCTTAACAAAAACCAGTAAGTGCTTA	2460
Qy	3167	tatgatgtgaatgtgatgatacaacattcatctaaacaggcaactgatgacaataaaaaa	3226
Db	2461	TATGATGTGAATGTGGATGATACAAACCATTCATCTTAACAGGCACCTGATGACAATAAAAA	2520
Qy	3227	cttggctcaaaacacccaactgaacaaaacaagtgtcaatggttaatacagcaactaac	3286
Db	2521	CTTGGGTCAAAACCCCAANCTGAACAAAACAAAGTGTAAATGGTAAATACAGCAACTAAC	2580
Qy	3287	tttaatgttaactctagtgatgaagatgcccttggtaacgccaaagacatccgcgaaat	3346

QY 2714 attaaatttactaatgtgaatgtagtaataatccaggtagctggaattgcaaaataacgctcgc 2773
 Db 44757 ----- 44756
 QY 2774 attaccagagataaaattggcttctgctggttctgagtggtgagtgatgatacaaaaacact 2833
 Db 44757 ----- 44756
 QY 2834 tatcttgatcaagacaagctacaaagtggcaatgttaagattaccaacacgtggcattaac 2893
 Db 44757 -----TATCGGTAAC 44766
 QY 2894 gcagggtggtaaaagccatcacagggctgtccccaacactgctcgtagcattggccga---tcaa 2950
 Db 44767 GATACCAAAAACATCATCAAGGGCTGTCTCAACACTGACTGACATTACCAACACAGAT 44826
 QY 2951 agtagcgcaacatagaactggcagatacaatactcaagacaaaagaataatccaaacgtgcc 3010
 Db 44827 AATACACACACCAAGCAAGCAATGACGCCCAAGGCAAG---AATCCAACGCTGCC 44883
 QY 3011 agcatlaatgatatataataacaggcttttaaccttaaaaaataataacaacccattgac 3070
 Db 44884 AGTATCAAGATGTGCTAAACGAGGCTTTAACTTGCAGGCAATGGCAAGCCCGTTGAC 44943
 QY 3071 ttgtctccacttatgacattgttgactttgccaattggcaattgccaccaccgcacagta 3130
 Db 44944 TTTGTCTCTACTTATGATACCGTCAACTTTTGCCAATGGCGATGGCACACAGTT 45003
 QY 3131 acc-----catgataccgctaacaacacagtaaaagtggatgatatatgagtgaattggat 3184
 Db 45004 ACTTAATTAATTAAGGCAACCAACACGACAGCTGTCTAATGTCATGTCATGTTGGAT 45063
 QY 3185 gatacaaccattcatctaaacagcactgatgacaataaaaaacttggcgtcaaaaacacc 3244
 Db 45064 GATACAACCATTCATCTGCCAATAATATGTTAATAAATAAAGCTTACGCTTAAACACC 45123
 QY 3245 aaactgaacaaaacagtgctaatggttaatacagcaactaaacttttaattgtaacttagt 3304
 Db 45124 ACATTGACCAAAAACAATGGTCAAAATGGTAATGCAACTAAATTTAGTGCCAAC----- 45177
 QY 3305 gatgaatgaccttgttaacgccaagacacatgcgcgaataataacacccctagccaag 3364
 Db 45178 AATGGGATGCCCTTGTTAACGCCAAGGTATTTGCCGCAANTTAACATCTTAGCCGAG 45237
 QY 3365 gaattcacaccacaaagcagcagacacacgcctcaaaaccttttacggttaaaaag 3424
 Db 45238 GAAATTCACACCACCAAGGCACAGCAGACACCGCCCTACAAACCTTTAAAGTCAAAAA 45297
 QY 3425 gtatgataaaaataaatgctgtagcagccaaacgcacatcccggtgggtcaaaaagacgca 3484
 Db 45298 GACGGTGCAACCAATGATGATG-----ACACCATCACCGTGGT----- 45336
 QY 3485 aataatcaagtcacaccccttaacactcaaaagtggaacggtcttaattataataaacccgac 3544
 Db 45337 AAAATGCCGTGCACACCCCTTAACATTAAGGTGAAGGTGAAGGTGAAGGTGAAGGTGAAG 45396
 QY 3545 aaaaatggtagcgttacctttggcattaaacacacagcggtcttaaaacggtg---caaa 3601
 Db 45397 AAGATGGGTACGGTTACCTTTGGCATTAACACCCCAAGCGGTCTTAAGCCGCGCACAC 45456
 QY 3602 agcaccttaaacacgctggttctgtctattataaaaaacccactggtgtagcgaacaaatccaa 3661
 Db 45457 ACCACTCTAAACAAGATGGCTGTCTATTAAAAACCCCACTGGTAACGAACAATCCAA 45516
 QY 3662 gtcggtgctgtaggggtgaattggcagagtttaataataatggtttagtgctggtgc 3721
 Db 45517 GTCGGTGTGATGGCGGTGAAGTTGGCAAGGTTAAATAAATGTTGTTAGTGTGCTGGC 45576
 QY 3722 attgagggacaactgcgattaccagagatgaattggcttttactgggactaatggtcga 3781
 Db 45577 ATTGATGGCAACAGCGGTATCACCACCAAGATAAAATTTGGCTTTCTGGGCTAATGGCTCA 45636

QY 3782 cttgataaaagcaaaacccacccctaagcaaaagcggcattaacgcagtggtgtaaaaaagatt 3841
 Db 45637 CTTGATACCACAAACCCACCTAAAGCAAAAGCGGATTAACGCAAGTGTGTAAGAAAGATT 45696
 QY 3842 accaaatcaatcagtgagattgccaaaacagccatgactgtagcagggcggaag 3901
 Db 45697 ACCAAATTCATCAGTGAGATTGCCAAAACAGCCATGATGCTGTGACAGGCGGTGAC 45756
 QY 3902 attatgattttaaaccgcaacttgaaacaaaatcagcagtagtactgcaaaaacagcaca 3961
 Db 45757 ATTTATGATTAGAAACCGAATTAAGAACAAATCAGCAGTACTGCCAANACGACAAA 45816
 QY 3962 aactattacacgaattctcagtagcagatgaacaaaggttaataactttacggttagtaac 4021
 Db 45817 AACTCATATACGAAATCTCAGTAGCAGATGAACACAGGTAAATAACTTTACGGTTAGTAAC 45876
 QY 4022 ccttactccagttatcacacctcaaaagacctcagatgctcatcccttgcaggtgaaaaac 4081
 Db 45877 CCTTACTCCAGTTATGACACCTCAAGACCTCTGATGTCATCACTTTGCGGTGAACAC 45936
 QY 4082 ggcattaccaccaaaggttaataaaggtgtgtgctgtgtggcattgaccaaaccacaaagc 4141
 Db 45937 GGCATTACCAACCAAGGTAAATAAAGGTGTGTGGTGGGCATTGACCAACCAAGGC 45996
 QY 4142 ttaacacgcctgaagctgacctgggtggttaataatggcgaagcattgtcattgacagc 4201
 Db 45997 TTAACACACGCTTAAGCTGACCGTGGGTAAATAATATGGCAAGGCATTTGTCATTGACAGT 46056
 QY 4202 caaaatgggtcaaaataccatcacaggactaagaacaaactagctaatgttaccatgat 4261
 Db 46057 AAGATGGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTTACCAATGA- 46115
 QY 4262 aaaggtagcgtacgcacccacagcaagggcaataataatcaaaagcagaacaaacccgt 4321
 Db 46116 -----TGGTCAGACACGCGACTAAGCCAAAGGCTTGCCGATACACCCGACAAACCCGT 46170
 QY 4322 gccgcagcactgttgatgtgctgaagcgcaggttttaacttgcaaggcaatggtggaagc 4381
 Db 46171 GCTGCCAGCATTTGGTGTATTAAATGCAAGGCTTTAACTTGAAGGCAATGGTACAGC 46230
 QY 4382 gttgactttgtccacttatgacacgttgcacattgcccagatggcgaatgcccacacgct 4441
 Db 46231 GTTGACTTTGTCTCCACTTATGACACTGTCACTTTGCCGATGGCAATGGCACACCGCT 46290
 QY 4442 aagtgaccttatgatgacacaaacaaacacagtaaaagtgtcttatgctgcaatgtggat 4501
 Db 46291 AAGTGACCTTATGATGACAGCAACCAACCAAGTCAAGTGTCTTATGATGTCATGTGGAT 46350
 QY 4502 gatacaaccattgaagttaaagataaaaaacttggcgttaaaaaacccacacattgaccagt 4561
 Db 46351 GATACAACCATTTGAAGTTAAAGATAAAAAACTTGGCGTAAAAACCCACCATTTGACCAGT 46410
 QY 4562 actggcacagtgcttaataaattggcctaagcaatacaagctactgaggtgctgtgc 4621
 Db 46411 ACTGGCACAGTGCTATAAATTTGCCCTAAGCAATCAAGCTACTGGCGATGGCGTTGTC 46470
 QY 4622 aagggcagtgatcgtgtctcatcctaaacaccttatctggcgacatcccaactgccaac 4681
 Db 46471 AAGGCCAGTGATTCGTTGCTCATCTAAACACCTTATCTGGCGNACATCCAAATGCCAAA 46530
 QY 4682 ggggcagcccaagcaaaactcagcaggtgtatgtgagtgctgtaggcaataaggtctac 4741
 Db 46531 GGGCAAGCCCAAGCAACAACTCAGCAGGCTATGTGGATGCTGATGGCAATAAGGTCATC 46590
 QY 4742 tatgacagtaccgatbaaagaagtactatcaagccaaaatgatggcacagttgataaaacc 4801
 Db 46591 TATCAGAGTACCGATACAAAGTACTATCAAGCCAAAAAATGATGSCACAGTTGATAAAC 46650
 QY 4802 aaagaagttgccaaaacacaaactggcgcgaagcccaaacccacagatggcattggct 4861
 Db 46651 AAAGAAGTTGCCAAGAACAACTTGGTCGCCCAACCCCAACCCGATGGCATTTGGCT 46710
 QY 4862 caaatgaatgtcaaatcagtcattatacaaaagaacagtaaatgatgccaataaaagcaa 4921

QY 2307 -----ttggcgtgaaacacccagcgttaacagtgatgcactagtgataaaatttagt 2359
DB 1753 GAACGACTTAAAGTGGGTGGTGTGAATATACCAGATAGTGGTATTAAATGCTGGTAAAT 1812
QY 2360 gttaaaggttagtgggtacgaacaattagcttagtaccgcgaacatttggcgaagctatcta 2419
DB 1813 CACAGATACCGGACTTACTAATGATAGCAATACCGATCGGTTTACCATCAACAG 1872
QY 2420 aatgaagtcaatcgaaacgctgacagtgctctacaaagctttaccggttaagaagaagac 2479
DB 1873 CTCAAGAGCGCAACGCTACTTTAAACGACGAGCGATGCATGATTAATAGTAATAAC 1932
QY 2480 gatgatgaac-----ccaaagcgtatcacccgtgct 2509
DB 1933 GGGGATCTAGTTAGTAGTGGCAATATATACCACCCCAACTTATAACATTTAGCGTGAA 1992
QY 2510 aaagatagacaaaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggt 2569
DB 1993 ACCACTAAGCTTAACAGTAATGGCACCAAGTGAATAATAATTTAGTTAGTAATGCT 2052
QY 2570 ctaaaggttgctaccaaagaagatggtacggtttacccttgggttagccaagatagcgt 2629
DB 2053 CATGATAACAATAGCTTAGTTACCGCCAAAGATTTGGCAGACTATCTAAATAAAGTCAAT 2112
QY 2630 ctgaccattggcaaaagcaccctcaaacacgatggtcttgactgttaaagataaccacgaa 2689
DB 2113 GAACCGCTGCAGTGCCTTACCAAGCTTAAAGTCCAAAACGGTGATTAATAGCAACAAC 2172
QY 2690 caaatccaaagtcgggtgctaatggcattaaattactaaatgtaagtggttagtaaccaggt 2749
DB 2173 GCCATCACCGTGGGTAAAGATCAAAACGCAAGACCTTCAACACCTTAAAACTCAAGGT 2232
QY 2750 actggcattgcaaataccctgcgacta---ccagagataaaattggcttgcgtgttct 2806
DB 2233 GAACCGGTGTTAATTAACGACCAATAGAGCCACAGGTACAGTTACCTTTGGCATTGAC 2292
QY 2807 gatgggtgcagttgatac-----aaacaaacttatcttgcataagcaagcta--caa 2857
DB 2293 CAAAGTAATGGTCTCACCGCTAAGCTGACCGTGGTGGGATACAAATGGTAATCGA 2352
QY 2858 gttggcaattgtaagattaccaactgcattaacgcagggtggtaaagccatcacagg 2917
DB 2353 TTGGTTATTGAGCAAGTCCCTAGCGCTGACGGTACAGCACCACCAATCATTAAGGA 2412
QY 2918 ctgtcccaacactgcctagctagctgcgcgaagtagccgaacataagaactgggcaat 2977
DB 2413 TTGTCCCAACACTGCTAGCATTTGCCAGTCCAAAGTGGCGCAACATAGCACTGGGCAAT 2472
QY 2978 acaatccaagacaaagacaaataccacgctgcccagattaaatgatataataatacaggc 3037
DB 2473 ACAATCGAAGAAAGACAAATCCAACTGCGCAGCATTTGATGATGCTGCTAAATGCGAGC 2532
QY 3038 tttaacctaaaaataatacaacccttgcattgtctccacttatgacattgttgac 3097
DB 2533 TTTTAACTTAAATAATTAATGCAAGACAAAGACTTTGTCTCCACTTATGACACTGTTGAC 2592
QY 3098 ttgccaatggcaatgccaccacccgacagtaaccatgatagctacaaacacaggt 3157
DB 2593 TTTATCGATGGCAATGCCACCCGACAGTAACCTATGATGAAGCCCAATCAACCACT 2652
QY 3158 aaagtggatatagtatgtaagtggatgatatacaaccattcatctaaacgacctgatgac 3217
DB 2653 AAAGTGGCGTATGATGTGAATGGATGGAGAAACCACTTGAATGACAGCGCGATAATGCG 2712
QY 3218 aataaaacttggcgtcaaaaccccaactgaacaaactgaacaaagtgctaatggtaataca 3277
DB 2713 AGAACAACCTTGGCTCAAAACCATCAACTGACCGCAACAGTACTAATGGTAATGCA 2772
QY 3278 gcaactaaactttaaagttaactctagtatgaagatgcccctgttgaacgccccaaagcatc 3337
DB 2773 AC-----TACATTTAGTACCGAGTACCATGACCTTGTGTTAAAGCCAGTGATATC 2823

QY 3338 gccgaaatctaaacacccctagccaaggaattcacaccaccaaaagcagcagacacc 3397
DB 2824 GCCGGCAATCTAAACACCCTAGCGAGGAAATTCACACCACCAAGGCACAGCAACACC 2883
QY 3398 gccctacaaccccttaacggttaaaaaaggttagatgaataatgctgatgacgccaac 3457
DB 2884 GCCCTACAAACCTTTACCGTTAAAGGTAGTGAATGATAAGGCTGATGACCCAAC 2943
QY 3458 gccatcccggtggtcacaagaacgcaataaacaagtccaacacccctaaacactcaaaaggt 3517
DB 2944 GCCATCACCGTGGTAAAGATGGCACAAGTGTAAAGTCAACACTTTAAACTCAAGGT 3003
QY 3518 gaaacggtcttaataataaaacgcgaacaaaatggtcaggttaccttggcattaaacac 3577
DB 3004 AAAACGCTCTGTGATTAATAAACCGCAAAAGATGCTACGTTACCTTTGGCATTTAACACC 3063
QY 3578 acaagcgtcttaaacgcgcgaacaaagcacc---ctaaacgcggtgctgtctctataaa 3634
DB 3064 CAAAGCGGTCTTAAAGCGCGGACAGCACCACTCTTAACAACAATGGCTTGTCTATTAA 3123
QY 3635 aaccccaactggtgacgaacaaaacccaagtcggtgctgagtcggtgaagtttgccaaggtt 3694
DB 3124 AACACCGCTAGTAACGAACAATAACCAAGTCGCTGTGATGCGTGAAGTTTGGCATGCT- 3182
QY 3695 aataataatggtgttagtgctggtggtggtggtggtggtggtggtggtggtggtggtggt 3754
DB 3183 --TAATAATGCTGTAGTGGTGTGGCATTTGATGCAAACTTCGCAAGTGTGCGTGGTGGT- 3240
QY 3755 attgcttactgggaactaatggtcactgtataaaagcaaacccccactaaagcaaacag 3814
DB 3241 ATTGGCTTTACTGGGACTTAATGGCTCACTTGTAAAGCAAAACCCCACTTAAGCAAGAC 3300
QY 3815 ggcattaaacgcaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 3874
DB 3301 GGCATTAAACGCGAGTGGTAAAGAAATTAACCAATTCATCAGGTGAGATTGCCAAAAAC 3360
QY 3875 agccatgctggtgacagcgcaagatttatattataaaacccgaacttgaaacacaaa 3934
DB 3361 AGCCATGATGCTGTACAGCGCGCAAGATTATGATTTAAACCCCAACTTTGAAATAA 3420
QY 3935 atcagcagctactgcgaacacagcaaaaactcaatctacacgaattctcagtagcagatgaa 3994
DB 3421 ATCAGCAGTACTGCCAAACACAGCACAAAATTCATTACAGGAATTCATAGTAGCAGTAA 3480
QY 3995 caaggtataactttacggttagtaaaccttactcactcactcactcactcactcactcactc 4054
DB 3481 CAAGGTAAATACTTTACCGTTAGTAACCTTACTCCAGTTATGACACCTTCAAGACCTCT 3540
QY 4055 gatgtcatcacctttgcaggtgaaacacgcaattacacacgaaggttaataaaaggttggtg 4114
DB 3541 GATGTCATCACCTTTGCAGGTGAAACGCAATTAACCAAGGTAAATAAGGTGCTGGTG 3600
QY 4115 cgtgtggcattgacaaaccaaaggttaacacgcgctaaagctgacggtggttaataat 4174
DB 3601 CGTGTGGCATTTGACCAAAACCAAGGCTTTAACACGCTTAAGCTAGCGTGGGTAAATAAT 3660
QY 4175 aatgcaagcattgtcattgacagcaaaaatggtcaaaaataccatcacaggaactaagc 4234
DB 3661 AATGCAAAAGGCATTGTCTAATACAGCCAAAATGGTCAAAATACCATCACAGACTAAGC 3720
QY 4235 aacactctagctaaatttaccatgataaaggttagctacgcacacacagcaagggcaat 4294
DB 3721 AACACTCTAGCTAATGTACTCAATGATAAGGTAGCTACGCCACACACAGGCAAT 3780
QY 4295 ataaataaagacgaagacaaaacccggtgcgcgacgattgttgatgctgtaagcagcaggc 4354
DB 3781 ATAATCAAAAGACGACAAAAACCCGTGCCGACGACATTTGTTGATGTGTAAGCGCAGGC 3840
QY 4355 tttaactgcaaggaatggtgaagcgttgacttgcctcacttatgacacccgccaac 4414
DB 3841 TTTTAACTTCAAGGCATAAGTGAAGCGGTGACTTTGCTCCACTTTATCACACCGCTCAAC 3900
QY 4415 ttgcccagtggtgcaatgcccaccacccgctgaaggtgacctatgtagacacagcaaaacaggt 4474

Db 3901 TTTGCCAATGCAATACCAACCGCTTAAGGTGACCTATGATGACACAAGCAAAACAGT 3960
Qy 4475 aaagtgtctatgatgcaatgtgtagatgatacaaacattgaagttaaagataaaaaactt 4534
Db 3961 AAAGTGTCTATGATGCAATGTGATGATACACCACTTGAAGTTAAAGATAAAAAACTT 4020
Qy 4535 ggcgtaaaaaccacacacattgaccagactgacagactgacagagtgctaaattgcccctaagc 4594
Db 4021 GGCCTAAAAACCAACACATTTGACCACTGTCGACAGTGTGCTTAATAATTTGCCCTAAGC 4080
Qy 4595 aatcaactactgcgcagctgctgcaagccagtaataatcgtgctcatctataaacacc 4654
Db 4081 AATCAAGCTACTGGCGATGCGCTTGTCAAGGCCAGTGATATCGTTGCTCATCTTAACACC 4140
Qy 4655 ttactgtgagacatccaaactgcaaaagggaagcaagcaagcaaacactgcagaggtctat 4714
Db 4141 TTATCTGGGACATCCAAACTGCCAAGGGCGCANGCCAACTCAGCAGGCTAT 4200
Qy 4715 gtgatgctgatgcaataaagtgcatctatgacagtagtaccgatacaagactcaagcc 4774
Db 4201 GTGGATGCTGATGCAATAAAGGTCTATGACAGTACCGATAACAAGTACTCAAGCC 4260
Qy 4775 aaaaatagccacagttgataaaacaaagaagttgccaaagcaaacactggtcgcccaa 4834
Db 4261 AAAAATGATGCAAGTTGATAAAACCAAGAAAGTTGCCAAAGACAACCTGGTCGCCCAA 4320
Qy 4835 gcccaaacccagatggcacattggctcaaatgaatgcaaatcagtcattcaacaaagaa 4894
Db 4321 GCCCAAAACCCAGATGGCACATTGGCTCAAAATGAATGTCAAATCAGTCATTAACAAAGAA 4380
Qy 4895 caagtgaatgatgcaataaaagcaagggcatcaatgaagcaaacgcctttgtttaaagga 4954
Db 4381 CAAGTAAATGATGCCAATAAAAGCAAGGCATCAATGAAGACAACCGCTTTGTTAAAGGA 4440
Qy 4955 ctgaaaagccgtctctgatacaaaaacaaacccagcagtaactgtggtgattta 5014
Db 4441 CTTGAAANAGCCGCTCTGATTAACAAACCAAAACCGCGCAGTAACCTGTGGGTGATTTA 4500
Qy 5015 aatgcggtgcccacacacccgtgacctttgcaggggatacagggcaacacggcttaaaaaa 5074
Db 4501 AATGCCGTTGCCAAACACCGCTGACCTTTGACAGGGATACAGGCACAAACGGCTAAAAAA 4560
Qy 5075 ctgggcagagacttgaccatacaagggtggaacacagacacacataagctaacgcgataat 5134
Db 4561 CTGGCGCAGACTTTGACCATCAAGAGTGGGCAAAACAGACACCAATAAGCTAACCGATAAT 4620
Qy 5135 aacatcggtgtgtagcaggtactgattgcttcaactgtcaaaactgcaaaagacctaaccc 5194
Db 4621 AACATCGGTGTGGTAGCAGGTACTGATGGCTTCACTGTCAAACTTGCCAAAGACCTAACCC 4680
Qy 5195 aatcttaacagcgttaattgcaggtggcaccacaaattgatgacaaaggcgtgctctttgta 5254
Db 4681 AATCTTAACCGCTTAATGCAAGTGGCACCACAAATGATGAANAAGGCATCTCTTTGTA 4740
Qy 5255 gactcaagcggctcaagccaaagcaaacaccctgtctaaatgccaatgagcgtggaactg 5314
Db 4741 CACGCAAAACGGCTAAGCCCAAGCAAAACACCCCTGTGTAAGTGGCCAAATGGGCTGGACCTG 4800
Qy 5315 ggtgcaaggtcatcagtaattggtgcaaggaagccacaaagataccgacgtgccaatgta 5374
Db 4801 GGTGGCAAGGTCACTAGTAATGTGGGCAAGGCACAAAGATACCGACGCTGCCAAATGTA 4860
Qy 5375 caacagttaaacgaagtacgcacactgttgggtcttggtaatgctggttaatgataaacgct 5434
Db 4861 CAACAGTTAAACGAAGTACGCAACTGTTGGTCT - - - - - TGGTAATGATAACGCT 4911
Qy 5435 gacggcaatcaggtaaacattgcgcatacaaaaagaccccaaatcgaagttcatctct 5494
Db 4912 CACGCAATCAGGTAACATTTGCCGACATCAAAAAGACCCCAAAATTCAGGTTCACTATCT 4971
Qy 5495 aaccgactgcatcaaaagcagcaggtacttggcggtaaaggttaataacgataccgaa 5554
|||||

Db 4972 AACCGCACTGTATCAAAAGCAGGACGGTACTTTGGCGGTAAAGGTAATAACGATACCCGA 5031
Qy 5555 aaacttgccactggtgtatatacaagtggcggtgataaaagcggcaacgcgtaaacgggat 5614
Db 5032 AAACCTGCCACTGTGTGGTGTACAAGTGGCGCTGGATAAAGACGGCAACGCCTAACGGCGAT 5091
Qy 5615 ttaagcaaatgttgggtcaaaaaccccaaaaagatggcagcaaaaaagccctgctcgcacat 5674
Db 5092 TTAAGCAATGTTGGGTCAAAACCCCAAAAAGATGCGACAAAAAAGCCCTGCTCGCCACT 5151
Qy 5675 tatacgcgcagctcagacacacattttgaccaaaccccgagaaagccatgacaga 5734
Db 5152 TATACGCCGCAGGTGACACCACTATGTGACCAACACCCCGCAGAAAGCCATTGACAGA 5211
Qy 5735 ataatgaacaaggtatccgcttcttccatgtcaacagatggcaatcaagagcctgtgta 5794
Db 5212 ATAAATGAACAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGTGA 5271
Qy 5795 caagggcgttaacgcatgactcaagtgcctcaggcaagcactcagtggcgataggtttc 5854
Db 5272 CAAGGGCGTAACGGCATTTGACTCAAGTGCCCTCAGGCAAGCACTCAGTGGCGATAGTTTC 5331
Qy 5855 cagccaagcagatggtgaagccgcgttgccatagtcagcaaacacccagcagcgaac 5914
Db 5332 CAGGCCAAGGCAGATGGTGAAGCCGCCGTTGCCATAGGCAGACAAACCCCAAGCAGGCAAC 5391
Qy 5915 caatccatcgccatcggtgataacgcacaaagccagggcgatcaatccatcgccatcggt 5974
Db 5392 CAATCCATCGCCATCGGTGATAAGGCACAAGCCACGGCGATCAATCCATCGCCATCGGT 5451
Qy 5975 acagggcaatggttagcaggtaagcactcctgggtgcactcggcgaccccaagcactgttaag 6034
Db 5452 ACAGGCAATGTTGAGCAGGTAAAGCACCTCTGTTGCCATCGCGACCCCAAGCACCTTTAAG 5511
Qy 6035 gctataacagcttacagtggtgtaataaacacagctttaccgattgcactcaaacagat 6094
Db 5512 CTTGATACAGTTACAGTGTGGGTAAATAACAAACCAAGTTTACCGATGCCACATCAAAACCGAT 5571
Qy 6095 gtcttgggtgggcaataaacatcacctgaccgaaagtaactcggttgccttaggttca 6154
Db 5572 GTCCTTGGTGTGGCAATAACATCACCTGACCGAAAGTAACCTCGTTGCTTGTAGTTCA 5631
Qy 6155 aactctgcactcagtgacggcacacacagcagggcacacaaagcaaaaaatctgcagggaca 6214
Db 5632 AACTCTGCCATCAGTGCAGGCACACACAGCAGGCACACAAAGCAAAAATCTGACGGCACA 5691
Qy 6215 gcaggtacaacacacacagcaggtgcaacccgttacggttaaaggcttgcctggacaacg 6274
Db 5692 GCAGGTACAAACCCACACAGAGGTGCCACAGGTACGGTTAAAGGCTTTGCTGGACAAACG 5751
Qy 6275 gcggtggtgcggtctccggtggtgcctcaggtgctgaacgcgtatccaaaaatgtgca 6334
Db 5752 GCGGTTGTCGGTCTCCGTCGGTGCCTCAGTGTGAACCGCTATCCAAAATGTGGCA 5811
Qy 6335 gcaggtgaggtcagtgccaccagcaccgatgcggtcaatggttagccagttgtacaagcc 6394
Db 5812 GCAGGTGAGGTGAGTGCACACAGCAGCAGCAGCAGTGTGTTAGCAGTTGTGTACAAAGCC 5871
Qy 6395 accaaagcattgccaacgcaacaaatgacttaccatgcttaccacaaacaaacaaat 6454
Db 5872 ACCAAGCATTTGCCAACGAACCAATGAGCTTGACCATGTATCCACAAACCAAAAT 5931
Qy 6455 aagccaatgcaggtatttcatcagcagtggtgcgtccatgcccacaagcctacatt 6514
Db 5932 AAAGCAATGACGGGATTTTCATCAGCGATGGGATGGCTCCATGCCACAAGCCTACATT 5991
Qy 6515 cctggcagatccatggttacccgggggtattgcccaccccaacaggtcaaggtgcggtgca 6574
Db 5992 CTTGSCAGATCCATGGTTACCGGGGTATTGCCACCCCAACAGGTCAAGGTGCGGTGGCA 6051
Qy 6575 gtggagctcgaaagctgcgataatggtcaatggttatttaaatcaatggttcaagcc 6634
Db 6052 GTGGGACTGTGGAAGCTGTCGGATAATGGTCAATGGGTATTATAAATCAATGGTTACGCC 6111
|||||

Qy	6311	gaacgcgctatccaaatgtgacgaggtgaggtcagtgccaccacgacccagtcgcggtc	6317
Db	2761	GAAGCGCGTATCCAAATGTGGCAGCAGGTGAGTGCAGTGCACACGACCGATTCGGTC	2820
Qy	6371	aatgtagcagttgtacaagccaccacaaagcattgccacgcaaccaatgaagctgac	6430
Db	2821	ATGGTAGCCAGTGTATCAAGCCACCCAAAGCATTGCCACGAACCAATGAGCTTGAC	2880
Qy	6431	catgctatccaccacaaacgaataaaggccaaatgagggatttcacagcagtcgcatg	6490
Db	2881	CATCGTATCACCACAAAGCAATAAGGCCAATGAGGGATTTCATCAGCATGCGCATG	2940
Qy	6491	gcgtccalgcacaaagcctacatctctggcagatccatggttacgcgggggtatgccacc	6550
Db	2941	CGGTCCATGCGACAAAGCCTACATTCCTGGCAGATGCATGGTTACCGGGGTATTGGCACC	3000
Qy	6551	cacaacgctcaagtgctggtggcagtggaagctgcgaagcgtgcgataaagtgccaatg	6610
Db	3001	CACAAACGGTCAAGTGGTGGCGATGGGACGTCTGAAGCTGTCTGGATTAATGGTCAATGG	3060
Qy	6611	gtatttaaaatcaatggttcagccgatacccaagggccatgtagggcgcgagttggtgca	6670
Db	3061	GTATTTAAATCAATGTTTCAGCCGATACCCAAAGGCCATGAGGGCGGCATGGTGCA	3120
Qy	6671	ggtttcacttttaa	6685
Db	3121	GGTTTTCACTTTTAA	3135
RESULT 10			
AX079922			
LOCUS	AX079922	6942 bp	DNA
DEFINITION	Sequence 10 from Patent WO0107619.		
ACCESSION	AX079922		
VERSION	AX079922.1 GI:13159441		
KEYWORDS	Moraxella catarrhalis.		
SOURCE	Moraxella catarrhalis		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.		
REFERENCE	1 (bases 1 to 6942)		
AUTHORS	Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.		
TITLE	Recombinant high molecular weight major outer membrane protein moraxella		
JOURNAL	Patent: WO 0107619-A 10 01-FEB-2001;		
FEATURES	CONNAUGHT LABORATORIES LIMITED (CA)		
source	Location/Qualifiers		
	1..6942		
	/organism="Moraxella catarrhalis"		
	/db_xref="taxon:480"		
BASE COUNT	2286 a 1621 c 1564 g 1471 t		
ORIGIN			
Query Match 37.1%; Score 2590; DB 10; Length 6942;			
Best Local Similarity 66.9%; Pred. No. 0;			
Matches 4802; Conservative 0; Mismatches 1105; Indels 1269; Gaps			
Qy	542	atgaatcacatctataaagtcattctttaacaagccacagcacattatggcagtgcca	601
Db	1	ATGAATCACATCTATAAAGTCATCTTTAAAGCCACAGGCACATTTATGGCGTGCA	60
Qy	602	gaglacgccaatcccacagcagcgggggggtagctgtctacagggcaagttggcagt	661
Db	61	CAGTGGCGCAAAATCCACACGCGGAGGATGACGATGATGACGAGGACAGGTGGGACG	120
Qy	662	gtatgcactctgagcttggcccgctattgcgcgctctgctctcgtatgcggtgcaacg	721
Db	121	TCTCCTGTCACTCCGCTGACTCGTGTGCCACGCTCGCTATCTAGTCATCGGTGGCAGC	180
Qy	722	ctcagtgccgagtgcttatgctcaaaaaaagatacacaacatcatcycaattggtgaaca	781
Db	181	CTCAATGGCAGTGTGTTATGCTCAA-----ATAATAGCAAGATCGCATTTGGT-----	228

QY	1856	agtagtactacagctgaattattgagtgatagtttaacaccttaccagcccaataacagc	1915
Db	1291	GGTAATAACACCGCTGAGCTACAAAGCGGTGGTTAACTTTACCC-----CAACAACA	1344
QY	1916	agtcacagcacaagcaaaacgcgtctatggcggttaattgggggtgaagtttactaataatgca	1975
Db	1345	AATGCAAGCACAGACAAACCGCTATGCGCAGCTGATGGGCTTAAGTTTACTGATAATCT	1404
QY	1976	gaacacacagcagcaatcgcactactcgtattaccagagataaaaattggcttctgcgcga	2035
Db	1405	AATACGGCACTTGAA---GATACCTACTCGTATACCAAGATAAAATGGTTTTAGCAAT	1461
QY	2036	gatggtgatgttgatgaacacacagccacattttgggataaaacacacttaaatgggt	2095
Db	1462	AAAGCTGGTACAGTTGATGAAACAAACCTTATCTTGTATAAGACAAGCTAAAAGTTGG-	1520
QY	2096	agtgttgcataatcaccatagacaaatggcagatgatgcaggttaataaaagatcagtaactt	2155
Db	1521	-----	1520
QY	2156	gccaaagtagcagtgctaacgtagcggttaccatcgaaacagctcaaaagcgcgaagcct	2215
Db	1521	--CAACAGCACCTTAACACACGGTGGCTTGACTGCTTAATAACACCATTTGGTGTAGCAAT	1578
QY	2216	actttaaacgcagcgctggcatcagtgctcacactactgaaatcagttgagtgcgtaag	2275
Db	1579	AAACAAATCCCAAGTCGGTGTGATGGCATTAATTTGCCGATGTAATTTAAATGTATFCA	1638
QY	2276	agtggcaatgttaacgcccccaacttaacacattggcgtaaaacacacagcagcttlaacagt	2335
Db	1639	AATGCCGGAATAATTCGGCACCTACTCGTATTAACCGAAGAGGAAATTTGGCTTTGCTGATGCT	1698
QY	2336	gatggcactagtgataaatttagtgttaagggttagtggtagcagaacaaatagcttagttacc	2395
Db	1699	GATGGTAAAGTTGATAA-----	1715
QY	2396	gccgaacattggcgaagctataatgaagtcaatcgacggcgctgacagtgctctacaa	2455
Db	1716	-----AAAGTACCATATTTGGTATAAANAACACTTCAAGTGGTGGTGT	1761
QY	2456	agctttacgcttaagaagaagacagatgatcagcgaacgcctatacccggtggctaaagat	2515
Db	1762	AAATTTACAAAGACAGCTGGCATTAATGCAGGTGATCAAAAGATCAGTAATGTTAAAGAT	1821
QY	2516	acgacaaaaaatgacggcgagtcagcatcttaaacctcaaaagtaaaaaacggtctaacg	2575
Db	1822	GCAACGGAGATACCGATGCGTGCATCTTATTAACAGCTTTAAACAGTCCACACAGCGCC	1881
QY	2576	gttgctacacaaaaagatggtagcgggttacctttgggcttagcgaagatagcggtctgacc	2635
Db	1882	GACGGTGGCTTACAA-----AGCTTCTCTATTTCGTGATGAAAMAGGTGAGGAA	1929
QY	2636	attggcaaaagcacccttaacacacagatggcttgactgttaaaagataaccaagacaaatc	2695
Db	1930	TTTACGATTAGTAACCTTGTATTTCTAATGTTAATACCCCAATACCTTTGAGACCATFACC	1989
QY	2696	caagtcggtgctaaatggcattaaatttactaatgtgaatggtagtaataccaggtactggc	2755
Db	1990	TTTGACGGTGAACCGCATCAGTATCAGCATGACATAGCCAAAGGTAAAGTCAAG--	2047
QY	2756	attgcaaataccgcgtcgcaattaccagagataaaaattggcttctgctggttctgattgga	2815
Db	2048	-TTGTATTGACCCCAATCAATGGTCTCACAGCGCTTAGCTGACCGTGGGTAGCGATAAA	2106
QY	2816	gttgatacaaaaaaaccttatcttgatcaagaacacagctacaagttggcgaattgaagatt	2875
Db	2107	GATGCTAAACATCAATTTGGTTATTGAGCAAG-----	2137
QY	2876	acacacactggcatlaacgcaggtggtaagccatcacagggtgtgtccccaacactgcct	2935
Db	2138	---TGCTTAGCGGTAAACGACACACCAACAAACATCATTTAGAGGATTGTCCCAACACTGCCT	2193
QY	2936	agcattggcgatcaaaagtgcgcaacacatagbaactgggcaataacataccaa---gac	2989

QY	782	aaccgccaagacgctcaggcactgccaagcggaagcgagcgatgacgagccattgctattggt	841
Db	229	---ACCACAGCAACAATGACAATGCCCTCGCTAGCAATGAAGCATCATTTGCTATTGGT	285
QY	842	gaacatgctaacgcacagcgcggttcaagccatcgccatcggttagtagtaa---taaaact	898
Db	286	AGTCCTTGAAGGCACATGCCAATCAAGCTATTGCTATCGGTGGTAGCAACACGATGCT	345
QY	899	gtcaatggaagcagtttggataaagtaggtaccgagtgctacggttcaagagtcacatgcc	958
Db	346	CGTAATCAAGCGCTTAATCAGAAGCGAGGTTCCACGCCAAGGTAAAGAGTCCATGCC	405
QY	959	atcggtggtgatgataaagcagtggtatgctcctcgatgctcagctgtagtgagctta	1018
Db	406	ATCGGTGGTGTATGCTGCTGAGGCTGATGCTCGATTGGCTTGGTGGTAGTGACATTA	465
QY	1019	catgttgctgacagcagtggaatcctaacaatccgaaggtactctgattaaagatctt	1078
Db	466	TATTTGGTAGGAATAGCACATACTTAATATCCAAATGGTCTTTC---TTAGCACTCTT	522
QY	1079	attaacgcccacagtcagttataaagaataacgaagctcaaaaggataatgatgataaata	1138
Db	523	ATTCAAAACCATACATGATTACGCCAAATACGAGACTCAAAATGGTCTCA---GAAATAT	579
QY	1139	agacgcacaacgcgaagcggaacacccagctactcagtgaggagccatgtcatatgcacag	1198
Db	580	AGACGCACAGCAGCAGAGGACACGCCAGTACTGACGTGGGAGCCATGGCATATGCANAAG	639
QY	1199	ggctatttttccaaacgcttggtaacggggcaacagctaaagtgccctattccctggca	1258
Db	640	GGTCAATTTGCCAACGCCCTTTGGTACACGGTCAACAGCTGAAGGCAATTAATTCCTTGCCA	699
QY	1259	gtaggtcttgcgcgaacacgcgagggccaatctacaacgctatggttctgatacaaca	1318
Db	700	GTAGGCTTTACCGCCCAAGCCGAAAGAGATATCAATCGCTATTGGTCTTAATGCACNA	759
QY	1319	tctagtcgttggagcgatagcccttggcgaggtactcgtcgcagctacagcagggcaggt	1378
Db	760	GCTATCAATATTGAGACACTAGCCCTTGGTGCAGATACCTCGAGTGTATTGGATTACGGT	819
QY	1379	attgcccctaggtaaggtctgttgcactcagagtgataataatcttagacggcctat	1438
Db	820	ATTGCCCTAGGTATGGTCTCAGATCCCTTAATAATAATAATAATAATAATAATAATA	876
QY	1439	acacaaaaacccagcgactagaccaccaagtctcaagccaccaataataacgaagcggt	1498
Db	877	GCTATGTACCAAGGTAATGGGTCAACACATAAATAATCGCTTAAAGCCACCGCAATGCT	936
QY	1499	ccactttccattggttagtaactctatacaacgtataaatacatcaatgctgggtgaggtgt	1558
Db	937	TTATTTTCCATTGGTAGTACATATCAAGCGTAAATCATCAATGCTCGGTGCAGGTAT	996
QY	1559	aataaaacgagtcgctcaatgtggcacagctagaagcggtggtgaagtggcctaaggag	1618
Db	997	GAGATACCCATCGGTCAATGTGGCACACCTTAAAGCGGTGAGAAATCTGGCTA---AG	1053
QY	1619	cgtagaattactttcagggtgtagataacagctactgaagcaaaaaatagtttggataat	1678
Db	1054	CGTCAAAATTTACTTTTAAAGGTGATGATAACGGTACTTGGGGTTTAAAGAAAAAATGGCGGAG	1113
QY	1679	actttaactataaaggtgtg---cagagacacacgcattacacgataataatcgt	1735
Db	1114	ACTTTAACCATTAAGGTGGTGGAGACCCAGCGGCAAGCTAACCGATATAATAACATTT	1173
QY	1736	gtggtataaagagcgtgataatagtggtctgaaggttaaaacttgctaaaaactttaacaat	1795
Db	1174	GGTGTGTAACAGATATAATACTGGTCTGAAAGTTTAAACTTCTGCTAAAAACCTTAAGCGGT	1233
QY	1796	cttactgaggtgataacacactaatataatgcccacacacacagattaaaggttagtagt	1855
Db	1234	CTTGAACACAGTTAGCACCAAAAAACCTTAACCGCCGAGGAGAAAGTTACGGTAG---GTAGT	1290

Db 2194 AGCAATTACCAATGAGTGGCGTACGACACACAGAAAGAGGCAATACAAATCACACAGCGAC 2253
Qy 2990 aaagcaaaatccaacgctgcagcattaaatgatatataataacacagcctttaacctaataa 3049
Db 2254 GAAGACAANAATCCAAGCGCGCAGTATCGGTGATATATAATAATACAGCGCTTTAACTAAAA 2313
Qy 3050 aataaacaacccccattgactttgtctccacttatgacatttggactttgccaatggc 3109
Db 2314 AATAATAGCAACTCCGTTGGCTTGTCTCCACTTATTAACACTGTGTGACTTATCGATGGC 2373
Qy 3110 aatgcaaccccgccacagtaaacccatgataccgctaaacaaaacccagtaagtggtatat 3169
Db 2374 AATGCCACCCCGCTAAGGTAACCTTACGATGAACACCAATCAAAACCCAGTAAAGTAACCTTAT 2433
Qy 3170 gatgtaattggaatgatacaaacattcatctcaacagcgcactgatgacaa---taaaaaa 3226
Db 2434 GATGTAATGGGATGAGAAAACCATTTGAACCTCACAGCGCATATGGCAAGACAAACAAA 2493
Qy 3227 ctggcgctcaaaaccccaactgaacaaacaaagtgcttaagtcttaagtatacagcaactaac 3286
Db 2494 ATTGGCGTCAAAAACCCACACACTGACCACAAACAAATGCTAATGGTAAAGCAAC----- 2546
Qy 3287 tttaatttaactctagtatgaagatgcccctgtttaacgccaagaacatcgccgaaat 3346
Db 2547 --CAACTTTTAGTACCAACCGGATAAGGATGCGCCTTGTAAACGCCAAAGACATCGCCGAAAT 2604
Qy 3347 ctaaacaccttagccaaggaattcacaccacaaagggcacagacagcacccgacctacaa 3406
Db 2605 CTAACACCCCTAGCAAGGAAATTCACACCACCAAGGACAGCAGACACCGCCCTACAA 2664
Qy 3407 accttaccgttaaaaggtagatgaataataatgctgatgacgcgaacgcgcacatcacc 3466
Db 2665 ACCTTTAAAGTCAAAAAGACGGTGCA-----ACTGATGACGAAACCATCACCGTGGGT 2718
Qy 3467 gtgggtcaaaagacgcaataataatcaagtcaaacaccctaacatcaaggtgtaaaacggt 3526
Db 2719 AAAGATGGTACACAAAACGGCAGACCGCTCAACACTCTAAACACTCAAAAGGTGA AAAACCGGT 2778
Qy 3527 ctttaattaaacccgcacaaaaatggacggttaccttttgccattaaacacacacagcggt 3586
Db 2779 CTAACGGTGTCTACCAATAAAGATGGTACGGTTACCTTTGGCATTAACACCCCAAGCGGT 2838
Qy 3587 cttaaacggcgcaaaagcaac---ctaaacgagcgttgctgtctattaaacacccacat 3643
Db 2839 CTTAAAGCGCGCAGACACCACTCTAAACAAGATGGCTGTCTATTAAAAAACCCCGCT 2898
Qy 3644 gtagcgaacaaatccaagtcggtgctgagtcggtgagtttgccaaaggttaataataat 3703
Db 2899 AGTAACGAACAAATCCCAAGTCGGTGTGATGGCGTGAAGTTTGCCAAAGGT---TGATAAG 2955
Qy 3704 ggtgtgtagtgctggtcattgagcacactgcgcattaccagagatgaaattggcttt 3763
Db 2956 GGTAAATCAAGCACTGGCATTTGATGGCACAAGCCGTATCACCACCAAGATCAAAATGGCTTT 3015
Qy 3764 actgggactaatggctcacttgataaagcaaacccccacacctaaagcaagac----- 3814
Db 3016 ACTGGGCTAATGGCTCAGTTGATACCAACCAACCAACCCACCTAACCAAGACAAGCTTAAA 3075
Qy 3815 -----ggcattaacgcaggtggttaaaagattaccaac 3847
Db 3076 GTGGGTGAAGTTGAATTAACCAACTGGCATTAACCGAGGTGGTAAAGAAATTTACCAAC 3135
Qy 3848 attcaatcaggtgagattgcccaaacagccatgctgtgacagcgccgaagatttat 3907
Db 3136 ATTCAATCAGGTGATATTACCAAAACAGCAATGATGCTGTGACAGCGGTGCGGTTTAT 3195
Qy 3908 gatttaaaacccgacttgaaacaaaatcagcagtagtactgcgaacacagcacaaaactca 3967
Db 3196 GATTTAAAAACCGAATTGAAAGCAAAATCAACAGTGTCTGTATAAACAGCACAAAACCTCA 3255
Qy 3968 ttacacgaattctcagtagcagatgaacaagggttaataactttacggttagtaacccctac 4027
|||||

Db 3256 TTACACGAATTCCTCAGTAGCAGATGAACAAGGTAATCACTTTACGGTTAGTAACCCCTTAC 3315
Qy 4028 tccagttatcacacctcaaaagacctctgatctcatcacccttcgagtgaaaaacgagcatt 4087
Db 3316 TCCAGTTATGACACTCAAGACCTCTGATGTCATCACTTTGCAAGTGA AAAAGCGGATTT 3375
Qy 4088 accaccaaggttaataaaggt 4147
Db 3376 ACCACCAAGGTAAATAAAGGTGTGTGTGGCATTTGAGCAATTAACCAACCCAAAGGCTTAACC 3435
Qy 4148 acgcctaagctgacccgtgggttaataataatggaagacagcattgtcattgacagcaaaaat 4207
Db 3436 ACGCCTAAGCTGACCGGTGGGTAAATAAATGCGCAAGGCAATTTGTCATTTGACAGTAAAGAT 3495
Qy 4208 ggtcaaaataaccatcacaggaactaagcaaacctctagctaatgttaccatgataaaggt 4267
Db 3496 GGTCANAATACCATCACAGGACTAAGCAACACTCTAGCTATGTTTACCAATGA-----T 3549
Qy 4268 agcgtacgcacccacagacaggggcaataataataatcaaaagcagaacaaaacccgtgcgcgc 4327
Db 3550 GGTGACGACACGCACTAAAGCAAGGGCTTGCCAAATGACACCGACAAAACCCGCTGCCGCC 3609
Qy 4328 agcattgttaagtcttaagcagcagcgttttaacttgaagcgaatggtgaagcgtttgac 4387
Db 3610 AGCATTGCTGATGTCTAAACGCGAGGGCTTTAACTTTGCAAGGCAATGTTGAAGCGGTTGAC 3669
Qy 4388 ttgtctccacttatgacaccgtcaactttgcgatggcaatgcccaccacgcctlaaggtg 4447
Db 3670 TTTGCTCCACTTATGACACTGTTGACTTTTATCGATGGCATGCCACCACCGCTTAAGGTG 3729
Qy 4448 acctatgatgacacaagaacaaacccagtaagtgctgtctatgtatgtcaatgtgtgatacaca 4507
Db 3730 ACCTATGATGACAAAGCAAAACCCAGTAAAGTGTCTATGATGTCAATGTGGATAATAAAA 3789
Qy 4508 accaattgaagttaa---gataaaaaacttggcgttaaaaaacccacacattgaccagtaact 4564
Db 3790 ACCAATTGAAGTGACAAGTGAATAAAAACTTGGCGTCAAAACCCACCACTGACCAAAAACA 3849
Qy 4565 ggacaggtgctataataattggcctaagcgaatcaagctactgctgcgatgctgtctgtcaag 4624
Db 3950 AGTCTAATGTTAGTGA---ACCAAAATTTAGTCCCGCGGATGGGATGCCCTTGTTTAAA 3906
Qy 4625 gccagtgatctgtgtctatctaacaacaccttatctgtggcgacatccaacactgccaaggg 4684
Db 3907 GCCAGTGATATCGCACCCATCTAAATACCTTGGCTGGCGACATCCAAACCCGCAAAAGGG 3966
Qy 4685 gcaagccaaggaacaactcaagcaggtctatgtgactgtgactggaataaaggtcatctat 4744
Db 3967 GCAAGCCAAGCAAGCAGCTCAGCAAGCTATGTGGATGCTGATGGCAACAAGGTCACTCTAT 4026
Qy 4745 gacagtcaccgatacaagctactatcaagcgaataatgatggcagctgtgataaaccaaaa 4804
Db 4027 GACAGTACCGGATAGAAGTACTATCAAGTCAATGACAGGGTCAAGTGGACACAAAACAAA 4086
Qy 4805 gaagttgccaaagacaactggtgcgcaagcccaaacccccagatggcacatttggctcaa 4864
Db 4087 GAAGTTGCCAAAGACAACACTGGTGGCCCAAGCCCAACCCAGATGGCAGATTTGGCTCAA 4146
Qy 4865 atgaatgtcaaatcagctattacaagaacaaagtaaatgatgccataaaaaagcaagggc 4924
Db 4147 ATGAATGTCAATCAGTCATTAACAAAGAGCAAGTAAATGATGTCCTAATAAAGCAAGGC 4206
Qy 4925 atcaatgaagacaacgcctttgttaaaggacttgaagagcgcctctctgataacaaaacc 4984
Db 4207 ATCAATGAGACACACGCTTTTATCAAGGGCTTGAANAACGCGCCCAAGACACCAAAACC 4266
Qy 4985 aaaaacccgcagtaactgt 5044
Db 4267 AAAAAACCGCAGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCGCTGACCTTT 4326
Qy 5045 gcaggggatacaggaacacacggttaaaaaactggcgagagactttgacatcaaaaggtgg 5104
Db 4327 GCAGGGGATACAGGCACAACGGCTAAAAAACTGGCGGAGACTTTTGACCATCAAAAGTGGG 4386
|||||

QY	5105	caaacagacacaaataagctaaccgataataacatcgtgtggttagcaggtactgtagc	5164
Db	4387	CAAAACAGACACAAATAAGCTAAACCGATAATAACATCGGTGGTGAAGCAGGTACTGATGGC	4446
QY	5165	ttcaactgtcaaaacttgcaaaagacctaaccatcttaacagcggttaatgcaggtgagcacc	5224
Db	4447	TTTCACTGTCAAACTTGGCCAAAGACCTTAACCAATCTTAACAGCGTTTAATGCAGGTGGCACC	4506
QY	5225	aaaattgatgacaaaggcgtgtcttttttagactcaagcggtaagccaaagcaaaacacc	5284
Db	4507	AGAAATGTATGAAAAAGGATCTCTTTTGTGTAGACGCAAAACGGTCAAGCCAAAGCAAAACACC	4566
QY	5285	cctgttactaaagtgcacaaatggcgtgactggtggc-----	5320
Db	4567	CCTGTGCTAAGTGGCCANTGGCTGGACCTGGGTGGCAAMCGCATCAGTAACATCGGTGCA	4626
QY	5321	-----	5320
Db	4627	GCTGTGTGATGATAACGATCGGTGAACCTTTTAAGCAGTTTAAATGAAGTTGCCAAAACGGTGC	4686
QY	5321	-----	5320
Db	4687	AACAACCTAAACAACAAAGTAACCTAGGTGGTGCATTAACCTTTGTGTGAACCGATGCC	4746
QY	5321	-----	5320
Db	4747	AATGGCAAGCCCATCAATGCCACCGATGCAAGCCCCCAAAAGCCCATCAAGGGCGCGAT	4806
QY	5321	-----	5320
Db	4807	GGTAAATACTATCACGCCAACGCCAACGGCGTACCTGTGGCAAAAGATGGCAAGCCCCATC	4866
QY	5321	-----	5320
Db	4867	ACCGATGGCGGACAAACTTGGCAATCTGGCAGCTCATGGCAAAACCCCTTGATCGAGGTCAAT	4926
QY	5321	-----	5320
Db	4927	CAAGTGTGGCAAGCCTAGCGCGCAACTCAGATGCCATCACCCCTAACCAACATCAAGTCC	4986
QY	5321	-----	5320
Db	4987	ACTTTGCCACAAATTGACACACAAACACAGAGTAATGCCAATGCCAGGGCAAGCCCCAAAGT	5046
QY	5321	-----	5320
Db	5047	CTGCCAGCCTATCAGCAGCACAGCAAAAGTAATGCTGCCAGTGTCAAAGATGTGCTAAAT	5106
QY	5321	-----	5320
Db	5107	GTAGGCTTTAACTTGACAGACCAATCAATCAAGTGGACTTTTGTCAAAGCCTATGATACC	5166
QY	5321	-----	5320
Db	5167	GTCAACTTTGTCAATGGTACAGGTGGCGACATCACAAAGCGTGGTAGTGTGTGATGGCAGC	5226
QY	5321	-----	5320
Db	5227	ATGAGTAACATCACCGTCAACACCGCCTTAGCAGCGACCGATGATGATGGCAATGTGCTT	5286
QY	5321	-----	5320
Db	5287	ATCAAGCCAAAGATGGTAAGTTCTACAAAGCAGACACCTCATGCCAAAGCGCTCACTA	5346
QY	5321	-----	5320
Db	5347	AAAGCAGCAAAATCAGCCAGTGTATGCCAAAACTCCAACCTGGTCTTAAGCCCTTGTTAACCCC	5406
QY	5321	-----	5320
Db	5407	AATGCTGGTAAAGCGCAGTACAGCGGATGCAGTGGCTCTTAATAACATTTATCAAAAGCGGTA	5466
QY	5321	-----	5320
Db	5467	TTTAAATCCAAAGATGTACAACTACTACACAGCTAAGCTCTGATGCATCAGTATCCAA	5526
QY	5321	-----	5326
Db	5527	GGCAAAGATAAACACACATCACCTTAAGCAAAGATGGGTGAATGTAGGGCGTAAAGTCTC	5586
QY	5327	atcagtaagtgtgggcaaaaggcacaacaaagataccgacgctgccaatgtacaacagcttaaac	5386
Db	5587	ATCAGCATGTGGTAAAGGCACAAAAGACACCGCTGCAATGTACAAACAGTTAAAC	5646
QY	5387	gaagtacgcaactgtgtgggtccttgggtaagtctgttaataatgataacgctgaacggaatcag	5446
Db	5647	GAAGTACGCAACTTGTGGGTCTTGGTAATGCTGTGAATGATAACGCTGACGGCAATCAG	5706
QY	5447	gtaaacattgcccagacatacaaaaagaccaaattcaggttcatcatctaaacgcacgtc	5506
Db	5707	GTAACATTTGCCGACATCAAAAAGACCCCAATTCAGGTTTCATCTTAACCGCAGCTGC	5766
QY	5507	atcaaaagcagcacggttacttggcggttaaaggtaataacgataccgaaaaaacttgccaact	5566
Db	5767	ATCAAAAGCAGCACGGTACTTTGGCGGTAAAGTAATAACGATACCGMAAAACTTGGCCACT	5826
QY	5567	ggtggtatacaagtggtggcgttgataaaagacggaacgctaacgagcgttttaagcaatgctt	5626
Db	5827	GGTGTGTACAAAGTGGCGGTGGATAAAGACGGCAACGCTAACGGCGATTTAAGCAATGTT	5886
QY	5627	tggttcaaaacccccaaaaagataggcagcaaaaagccctgctgcgcaacttataacgcgcga	5686
Db	5887	TGGGTCAAAACCCAAAGATGGCAGCAAAAGCCCTGCTCGCCACTTATAACGCCCA	5946
QY	5687	ggtcagacaaactattgacaaacacccccagagaagccattgacagataaaatgaacaa	5746
Db	5947	GGTCAGACCAACTATTGTACCAACAAACCCCGCAGAAAGCATTTACAGAGTAATAATGAACAA	6006
QY	5747	ggtatccgctcttccatgtcaacgatggcaatcaagagcctgtgtcaagggcgtaac	5806
Db	6007	GGTATCCGCTTTCCATGTCAACGNTGGCAATCAAGAGCCTGTGTACNAGGGCGTAAC	6066
QY	5807	ggcattgactcaagtgcctcagggcaagcactcagtggtcgataggtttccagggccaagca	5866
Db	6067	GGCAITGTACTCAAGTGCCTCAGGCAAGCACTCAGTGGCGATAGGTTTCCAGGCCAAGCA	6126
QY	5867	gatgtgaagccgcgtgtgccaatagcagacaaaccccaagcagcaacaaatccatccgc	5926
Db	6127	GATGTGAAGCGCGCTTGCCATAGCGACAAACCCCAAGCAACCCCAAGCAACCAATCCATCGCC	6186
QY	5927	atcgggtgaacgcgcaagccacggtcgatcaatccatcgccatcggttacaggccaatgtg	5986
Db	6187	ATCGGTGATACGCACAAAGCCACGGCGATCAATCCATCGCCATCGGTACAGGCAATGTG	6246
QY	5987	gtagcaggtaagcactctgtgtccatcggcagcccaagcactgttaaggctataacagt	6046
Db	6247	GTAACAGGTAAAGCACTCTGTGCCATCGCGCACCCCAAGCACTGTAAAGGCTGATTAACAGT	6306
QY	6047	tacagtgtgggtataaacaacacagtttaccgatgccactcaaacccgatgtctttgggtg	6106
Db	6307	TACAGTGTGGGTAAATAACAACCAAGTTTATCGATGCCACTCAGACCGATGCTCTTGGTGTG	6366
QY	6107	ggcaataacacatcacccgtgaccggaagtaactcgggttgcctttaggttcaaacctctgcac	6166
Db	6367	GGCAATAAATACCTACCGTACCGAAAGTAACCTGGTTGGCTTAGGTTCAAACTCTGCGCATC	6426
QY	6167	agtgcaagcacacacgagcacacacaaacaaatctlgacggcacagcaggtacaacc	6226
Db	6427	AGTGCAGGACACACCCGACGACACAAAGCCAAATAATCTGACGGCACACAGGTTACACC	6486
QY	6227	accacagcaggtgcaacccggtacggttaaaggccttgcctggacaaacgcggttgggtg	6286
Db	6487	ACCACAGCAGGTGCAACAGGTACGGTTAAAGGCTTTGCTGGACAAACGGCGTGGTGGG	6546
QY	6287	gtctccgtgggtgcctcaggtgctgaacgcggtatccaaaatgtggcagcaggtgaggctc	6346


```
Db 6547 GTCTCCGTGGTGCTCAGGTGCTGAACGCGGTATCCAAATGTGGCAGCAGGTGAGGTC 6606
QY 6347 agtgaccaccagccgatgcggtcaatggtagccaggtgtacaaaagccacccaaagcatt 6406
Db 6607 AGTGGCCACAGCAGCCGATGGGTCAATGGTAGCCAGGTGTGTACAAAGCCACCCAGGCATT 6666
QY 6407 gcaacgcgaaccaatgagcttgaccatcgatccacacaaagaaataaggcgaatgca 6466
Db 6667 GCCAACCAACCAATGAGCTTGACCATCGTATCCACCAAGCAAAATAAAGCCAAATGCA 6726
QY 6467 ggaatttcacagcatggcgatggcgatccatgcccaagccctacattcctgcagatcc 6526
Db 6727 GGGATTTCATCAGCATGGCGATGGGTCTCATGCCAAGCCATACATCTCTGGCAGATCC 6786
QY 6527 atggttaccgggggtattggccaccacacagcgtcaagtgctggtggcagtggtgctcg 6586
Db 6787 ATGTTTACCAGGGGATTTGCCACCCACACAGCGTCAAGTGGCGGTGGCAGTGGGACTGTGC 6846
QY 6587 aagctgcgataatggtcaatgggtatttaaatcaatggttcagccgatacccaagc 6646
Db 6847 AAGCTGTGCGATATGTCATATGGTATTTTAAATCAATGGTTTCAGCCGATACCCAAAGC 6906
QY 6647 catgaaggcgagctgtgctgaggttttcacttt 6682
Db 6907 CATGTAGGGCGGAGTGTGTGAGGTTTTCACTTT 6942

RESULT 11
AX079959 AX079959 2448 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 47 from Patent WO0107619.
ACCESSION AX079959
VERSION AX079959.1 GI:13159466
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 2448)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of
Moraxella
JOURNAL Patent: WO 0107619-A 47 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
    source
        1..2448
            /organism="Moraxella catarrhalis"
            /db_xref="taxon:480"
BASE COUNT 826 a 569 c 556 g 497 t
ORIGIN

Query Match 35.1%; Score 2448; DB 10; Length 2448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3551 ggtacggttaccttggcattacacacacagcggtcttaagccggcgaagcaccta 3610
Db 1 GGTACGGTTACCTTGGCATTACACCAAGCGGCTTTAAGCGCGCAAGCACCCCTA 60
QY 3611 aacgacggttgcttgcattataaaacccactggtgagcgaacaaatccaaagtcggtgc 3670
Db 61 AACGACGGTGGCTGTCTATTAAAAACCCACCTGGTAGCGAACAATAATCCAAAGTCGGTGTCT 120
QY 3671 gatggcggaagttggcaagggttaataatggtgtgttagtgctggcattgatggc 3730
Db 121 GATGGCGTGAAGTTGGCAAGGTTAATAATGTTGTTAGTGTGCTGGCATTGATGGC 180
QY 3731 acaactgcattaccagagatgaattggctttactgggactaatggtcacttgataaa 3790
Db 181 ACAACTGCGATTACCAGAGATGAATTTGGCTTTACTGGGACTAATGGCTCATTGTATANA 240
```

```
QY 3791 agcaaaccccccaactaagaagaagcggcattaaagcaggttgtaaaaaagattaccaacatt 3850
Db 241 AGCAAAACCCACCTTAAGCAAAAGACGGCATTAACGACGTTGTTAAAAAGATTACCAACATT 300
QY 3851 caatcaggtgagattggccaaaacagccatgctgctgacagcggaagatttatgat 3910
Db 301 CAATCAGGTGAGATTGGCCAAAACAGCCATGCTGCTGACAGCGGCAAGATTATATGAT 360
QY 3911 ttaaaacccgaacttgaaaaacaaatcagcagctactgccaacaaacagcacaaaactcatta 3970
Db 361 TTAATAAACCCGAACCTTGAAACAAAAATCAGCAGTACTGCCAAAACAGCACAAAACCTATTA 420
QY 3971 cagcaattctcagtagcagatgaacaaaggtaataacttttacggttagtaacccttactcc 4030
Db 421 CACCAATTTCTCAGTAGCAGATGAACAAGGTAAATAAATTTACGGTTAGTAACCCCTTACTCC 480
QY 4031 agttatgacacctcaagacctctgatgcatcacctttgcagctgaaacggcattacc 4090
Db 481 AGTTATGACACCTCAAGACCTCTGATGTCTACCTTTGCAGGTGAANAACGGCATTTACC 540
QY 4091 accaaggtaaataaagtggtgctggtggtggtggtggtggtggtggtggtggtggtggtggt 4150
Db 541 ACCAAGGTAAATAAAGGTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600
QY 4151 cctaagctgacctgggtgataataataatggaagagcattgtcattgacagccaaaatggt 4210
Db 601 CCTAAGCTGACCGTGGGTAAATAATATGGAAGGCATTTGTCAATTGACAGCAAAATGCT 660
QY 4211 caaaataccatcagcagtaagaacaacactcagctaatgtttaccaatgataaggtagc 4270
Db 661 CAAATACCATCAGGACTAAGCAACACTCTAGCTAATGTGTACCAATGATAAAGGTAGC 720
QY 4271 gtacgcaccacagaacaggcgcaataataataaagaacgaacacaaacccgtgcccgcagc 4330
Db 721 GTACGCACACAGACAGGCGCAATATAATCAAGACCAAGACAAACCCGTGCCGCCAGC 780
QY 4331 attgttgatgctgaagcaggcgttttaacttgcaaggcaatggtgaagcggttgacttt 4390
Db 781 ATTGTTGATGTCTAAGCGCAGGCTTTAACTTGAAGGCAATGTTGAAGCGTTGACTTT 840
QY 4391 gtctccacttatgacacggtcaacttgcgagtggaatgccaccacgctaaagtgaac 4450
Db 841 GTCTCCACTTATGACACCGTCAACTTTGCCGATGGCAATGCCACCGCTTAAGGTGACC 900
QY 4451 tatgatgacacgaacaaacacagtaaaagtgctctatgatgtcaatggtgatacacaac 4510
Db 901 TATGATCACACACCAACCAACAGTAAAGTGTCTATGATGTCAATGTGGATGATACAAC 960
QY 4511 attgaagttaagataaaaaacttggcgttaaaacccaccacattgaccagttactggcaca 4570
Db 961 ATTTGAAGTTAAAGATAAAAAAATTTGGCGTAAAAACCCACCATTTGACCACTTACGACCA 1020
QY 4571 ggtgctaataaatttgcctaagaacaatcaagctactgagcgtgctgctgcaagccagt 4630
Db 1021 GGTGCTAATAAATTGGCCCTAAGCAATCAAGCTACTGGCGATGGCTGTGTCAAGGCCAGT 1080
QY 4631 gatcgtgtgctctatcaaacaccttatctggcgacatccaaactgcaaaagggcgcaagc 4690
Db 1081 GATATCGTTGCTCATCTAAACACCTTATCTGGCGACATCCAAATGCCAAAGGGGCAAGC 1140
QY 4691 caagcgaaactcagcagcgctatgtggtgctgagtgagcaataaggctcatctatgacagt 4750
Db 1141 CAAGCGAACAACCTCAGCAGGCTATGTGGATGTGTGATGCAATAAGGTCATCTATGACAGT 1200
QY 4751 accgataaagttactatcaagccaaaatgatggcacagttgtataaaacaaaagaagtt 4810
Db 1201 ACCGATAAAGTACTATCAAGCCAAAATGATGGCACAGTTGATAAACCAAAAGAGATT 1260
QY 4811 gccaaagacaaactgctgcggccaaagccaaacccagatggcacattggctcgaatgaat 4870
Db 1261 GCCAAACACAAACTGGTGGCCCAAGCCCAACCCAGATGGCACATTTGGTCAATGAAT 1320
QY 4871 gtcaaatcagtcattatacaagaagaagaatgataatgataaaagaagcaggtcatcaat 4930
```



```
RESULT 13
AE006194/c
LOCUS
DEFINITION
    AE006194 10533 bp DNA BCT 08-MAR-2001
    Pasteurella multocida PM70 section 161 of 204 of the complete
    genome.
ACCESSION
    AE006194 AE004439
VERSION
    AE006194.1 GI:12721958
KEYWORDS
    .
SOURCE
    Pasteurella multocida.
    Pasteurella multocida.
    Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
    Pasteurella.
    1 (bases 1 to 10533)
    May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
    Kapur,V.
    Complete genomic sequence of Pasteurella multocida, Pm70
    Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
    21145866
REFERENCE
    2 (bases 1 to 10533)
    Zhang,Q. and Kapur,V.
    Direct Submission
    Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
    University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
    55108, USA
FEATURES
    source
        1..10533
        /organism="Pasteurella multocida"
        /strain="PM70"
        /db_xref="taxon:747"
        91..804
        /gene="PM1566"
        91..804
        /gene="PM1566"
        /codon_start=1
        /transl_table=11
        /product="unknown"
        /protein_id="AAK03650.1"
        /db_xref="GI:12721959"
        /translation="MTKLHLAWIFLALPIALNAQPDNDKLNQVQSFSAEVKFEIKDKD
        ELQVTLYLQEBGDASALNLIVARMNQALDPAVKKHVATVQVYSQQRVTOVRYGKEGQ
        NGWIDRAEIVLKSQDPMPLSKVLNEHFPIGGMYATVSEESLARVEKMTFAVLQK
        FEQKALQSLMKAGKRYLIELDSPHDVNGSTRSVAVSMKSKDAFYSSADHEMALE
        GKTKLRVYQAKIELND"
        868..1467
        /gene="fkpA"
        /note="PM1567"
        868..1467
        /gene="fkpA"
        /codon_start=1
        /transl_table=11
        /product="fkpA"
        /protein_id="AAK03651.1"
        /db_xref="GI:12721960"
        /translation="MKIAKNVVSIAQVYRTEDGVLVDEAPVNPQPLEYLGQHNNLVIG
        LENALGKAVGDFEVYRPEEAIGEYNENMVORPKDVFQGVDELVGVGMRFITADYDI
        GPLPVVITEVAGNDVVVDGNHMLAGQELLFSVEVATREATLEEIAHGHIEGGGCGC
        GHHDDEHGCGCGSHHHEHHEHHAHDGCGGCGCKH"
        complement(1521..1733)
        /gene="PM1568"
        complement(1521..1733)
        /gene="PM1568"
        /codon_start=1
        /transl_table=11
        /product="unknown"
        /protein_id="AAK03652.1"
        /db_xref="GI:12721961"
        /translation="MPEFRDLISKLEDAHNRLFEKHNELDHRIKNIENIELGT
        POEVNLKKEKLKDEIYVILKKH"
        complement(1830..2300)
        /gene="yfkJ"
        /note="PM1569"
        complement(1830..2300)
        /gene="yfkJ"
        /codon_start=1
        /transl_table=11
        /product="yfkJ"
        /protein_id="AAK03653.1"
        /db_xref="GI:12721962"
        /translation="MTIHLFLVCLGNICRSPMAEFLMRKIKOAGLETIVTKTSAGTS
        GMDHGMCGTADIIQDCHIDSTGFRSKRVQHQWQEFYIIAMDNSNLDLRLFG
        HHPDKLFKITELCPDILGYPWYSKDFNPTQYLLNPCCDALLARLEKHL"
        complement(2415..6314)
        /gene="hsf_2"
        /note="PM1570"
        complement(2415..6314)
        /gene="hsf_2"
        /codon_start=1
        /transl_table=11
        /product="hsf"
        /protein_id="AAK03654.1"
        /db_xref="GI:12721963"
        /translation="MNIYITLWNAATQSVVVSSELAAGKGSASKSALVNSVSGFS
        FTLIAASVILGSGVNAAEVTGNTGVSDDKYCFYNASSQSVICGDAITTKTDKNK
        PAKSVVIGCATNDGETNVAIGAKSSKSAASIAIGDNKALDNQAIAGONATANS
        WDISIGRGAQEOTEVSAEGRNIAIGDGALRGKGVNNIALCTSGADRAGTHNVLM
        GTVNADEAVRSALTAGSTTKEINAKETTDNKYIEASNTVALGTALATQLAAVIG
        QQAKFGNSVAVNGTKASGQTATGTSNAHATGSSAIGTIVNGDNTTTLASDN
        FTIAMLSAQATKSDAIAIGRNAKAGINTVSGINAGVTGTADEAQSDDPNKLEPAT
        DAVEIGNKAGYKSNRMQVSLGDSGEGVYGTENTVIGNSAGNTKGNNTVAISRA
        CONVEGHDFNFAALIEAGQNTIAIGHAGRSADPTKLNINNTISLKGESVSLK
        NFGIAOGRKAKTDGLASIAIGRNAEAVGGTANIAIGDSASADASGAIVLGTAKAQS
        LTVDGKKYGAISAIVIGTEAKIAQAAPAGKNPKDAIAIGTAKAHHAYASTALGFG
        AKSDTKAQAIVSICGNSKAGYQAIAGFSEAKTTENAGSSITAFGTQATRASASTIGM
        GAETGDFGQALGSDAVALGREAQKQNALAGYKADHDKDALVALGAGAAETAAAE
        GTNEATVNEFKYSGFAGIKPATVSVGKKDAERTITNVAAGRIDKTSTDAINGSQLYL
        ALNALGVNGLTVNLVNGDAATVKGDEAGTLMNSIGTGKGTTHDAIEAVNNTAK
        AKSTTVKEGDNITVTEAADSGRTVTATKDVKDFSVVAGCTKIDANGLTFVDDQG
        TKIDNTPSISKIDAGNKVTVNQNGNTAKDSKDAVNGQQLPAQGEYKNIIGDGT
        YNPTEAINTNIGGTGASTIDEIKAVNTTAKAEVTVQGENIVVTSAPGANGTV
        YTVATAKEVTFDKTTSVVVDITNDITGLSNKTLGGDNFNAKNGRAASBEQLNATQT
        NLATLVGNAQNTNGNVMDTIGTGKNNINDAIKASREVKQGNKMMVPTTGANGQ
        TTYEVATADKAVDFEVKVGITIDATNTKISGAKGDISENSTDAVNGSOLYELQOKI
        AKGSDNINLNNKDKDLRAGIAGANAAGLPQAYIPGKSMVAVAAGTYKQGNAI
        ALGMSRISDNGKVIILKLTGNTNSRGDEGASIGAGYQW"
        complement(6493..7833)
        /gene="PM1571"
        complement(6493..7833)
        /gene="PM1571"
        /codon_start=1
        /transl_table=11
        /product="unknown"
        /protein_id="AAK03655.1"
        /db_xref="GI:12721964"
        /translation="MTSSLPNIGFISLGPKNLVDSEIRILTELRSDGYNIPSVENAD
        LVIVNTCGFIDSVAQESLEAIGBALENGKVIIVTCLGAKEDRIREVHPKVLVETGPH
        SYEAVNQVHKYVPKPAYNPVNLVPQGVKLPKHAYALKISEGCDHRCFTCLIPSM
        RGDLSRSITQVLDEAKRLVEAGVKEILVVSQDTSAYALDRSEKQNTVFWNGMPKIK
        NNLSICEQVGLKGLVWVHYVYPHVDQLIPMAEGEKILPYLDLIPLOHASPKILKA
        MKPSIEFTLERIKKREICPDLTIRSTIVGFPGESEEDFQLLDLFLKALLDRVG
        CFKFSVEGAIAATEMPQVPEVEKARFHRFMOLQOEISAAALQOKVKVFTVLVDEV
        DEEGITARSIAIDAPEIDGVVYIDNPNRVAVKAGQFIEVKITRADAYDLASLIN"
        complement(8029..8787)
        /gene="udp"
        /note="PM1572"
        complement(8029..8787)
        /gene="udp"
        /codon_start=1
        /transl_table=11
        /product="udp"
        /protein_id="AAK03656.1"
        /db_xref="GI:12721965"
        /translation="MSEVPHGLTKAMLEGATLAIIPGAPERVEKIAKLDLRPKFLAS
        TREFTSWLGVIGDKAVVVCSTGIGGSPVSIATVEAQLGVRTFLRTGTCAGIQPHIN
        GDILVTTGAVRLDGASLHFAPMEYPAVANFECTNALLYKAATELANOKVYVGTAAADT
        FYPGQERYDTYSGKVYHFHGSLSKQMODLNMFMESATLFTMCSALGRAGWVAGA
        IVNRTOOEIPNEAAVKDIEKNAVEIVVKAAYLL"
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:38 ; Search time 1369.14 Seconds
(without alignments)
3197.431 Million cell updates/sec

Title: US-09-361-619-5

Perfect score: 6972

Sequence: 1 ccattgatattggcagggtgt.....ccgttgacctgcagggtgac 6972

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	6972	100.0	6972	22	AAF59102 M. catarrhalis str
2	6961	99.8	6973	22	AAF59100 M. catarrhalis str
3	6959	99.8	6971	17	AAT38740 Moraxella outer me
4	6258	89.8	6259	22	AAF59106 M. catarrhalis M56
5	6144	88.1	6144	22	AAF59103 M. catarrhalis str
6	5979	85.8	5979	22	AAF59101 M. catarrhalis str
7	4671.2	67.0	62909	22	AAF28545 Genomic fragment #
8	4101.6	58.8	6159	22	AAF59104 M. catarrhalis str
9	2590	37.1	6942	22	AAF59105 M. catarrhalis les
10	720	10.3	720	22	AAF59129 M. catarrhalis str
11	426.8	6.1	1000	22	AAF91426 Moraxella catarrha

12	148	2.1	3030	21	AA92496	Haemophilus influe
13	146.4	2.1	3036	21	AA92493	Haemophilus influe
14	141.6	2.0	3354	21	AA92495	Haemophilus influe
15	131.2	1.9	3342	21	AA92498	Haemophilus influe
16	125.2	1.8	936	22	AAF58252	Oligonucleotide D1
17	125.2	1.8	936	22	AAF58254	Oligonucleotide D1
18	125.2	1.8	936	22	AAF58257	Oligonucleotide D1
19	125.2	1.8	936	22	AAF58259	Oligonucleotide D2
20	125.2	1.8	936	22	AAF58262	Oligonucleotide D2
21	125.2	1.8	936	22	AAF58255	Oligonucleotide D1
22	124.8	1.8	936	22	AAF58252	Oligonucleotide D1
23	124.8	1.8	936	22	AAF58254	Oligonucleotide D1
24	124.8	1.8	936	22	AAF58257	Oligonucleotide D1
25	124.8	1.8	936	22	AAF58259	Oligonucleotide D2
26	124.8	1.8	936	22	AAF58262	Oligonucleotide D2
27	124.8	1.8	938	22	AAF58255	Oligonucleotide D1
28	85.4	1.2	7486	19	AAV22837	Haemophilus para
29	83.8	1.2	2079	21	AA92494	Haemophilus influe
30	83.8	1.2	8930	19	AAV22834	Haemophilus para
31	83.6	1.2	2037	17	AAT41477	Haemophilus adhesi
32	82.8	1.2	3294	17	AAT41475	Haemophilus adhesi
33	78.2	1.1	3300	21	AA92497	Haemophilus influe
34	78.2	1.1	7253	21	AA92499	Haemophilus influe
35	78.2	1.1	7291	17	AAT41476	Haemophilus adhesi
36	76.8	1.1	1830121	17	AAT42063	Haemophilus influe
37	67	1.0	1757	21	AA92500	Haemophilus influe
38	61	0.9	2308	20	AA85788	DNA encoding a sur
39	60.2	0.9	1797	20	AA85794	DNA encoding a sur
40	60	0.9	60	17	AAT38741	Moraxella outer me
41	59.4	0.9	102634	21	AA81464	N. meningitidis pa
42	59.4	0.9	349980	21	AAF21609	Neisseria meningit
43	59.4	0.9	1437668	21	AAA81490	N. meningitidis B
44	58.8	0.8	244	22	AAF58238	Oligonucleotide D1
45	58.6	0.8	1797	20	AA85790	DNA encoding a sur

ALIGNMENTS

RESULT 1

AAF59102
ID AAF59102 standard; DNA: 6972 BP.

XX AAF59102;

XX 24-APR-2001 (first entry)

XX M. catarrhalis strain 4223 genomic 200kDa gene SEQ ID NO:5.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection; ds.

XX Moraxella catarrhalis.

XX WO200107619-Al.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX P-PSDB; AAB69134.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis

XX

Db	1801	 tggggtgaa	tacacacacat	taaa	tgcacacacacag	ttaaggtaggtagtagtag	1860															
Qy	1861	 tactcacagctga	atattg	agtgatag	ttaaccttacc	cagccccaatcacaggcagtc	1920															
Db	1861	 tactcacagctga	atattg	agtgatag	ttaaccttacc	cagccccaatcacaggcagtc	1920															
Qy	1921	 aagcacaaag	caaaac	cgtctat	ggcgtta	atgggggtgaagtttactataatgcagaac	1980															
Db	1921	 aagcacaaag	caaaac	cgtctat	ggcgtta	atgggggtgaagtttactataatgcagaac	1980															
Qy	1981	 aacagcagca	atcggcact	actcgratt	taccagagata	aaaaattggcttggctcgagatgg	2040															
Db	1981	 aacagcagca	atcggcact	actcgratt	taccagagata	aaaaattggcttggctcgagatgg	2040															
Qy	2041	 tgatgttgatg	aaaaa	acagacc	attttggata	aaaaaacaacttaagtggtgtagtgt	2100															
Db	2041	 tgatgttgatg	aaaaa	acagacc	attttggata	aaaaaacaacttaagtggtgtagtgt	2100															
Qy	2101	 tgcgaattacc	atagaca	atggcatt	gtagcag	gttaataaaagatcagtaattcttgccaa	2160															
Db	2101	 tgcgaattacc	atagaca	atggcatt	gtagcag	gttaataaaagatcagtaattcttgccaa	2160															
Qy	2161	 aggtagcag	tgttaac	gatacg	gtttac	atcgcagagctcaaaagcccaagcctacttt	2220															
Db	2161	 aggtagcag	tgttaac	gatacg	gtttac	atcgcagagctcaaaagcccaagcctacttt	2220															
Qy	2221	 aaacgcagc	gcgtgc	atcag	gttcac	atcgcagagctcaaaagcccaagcctacttt	2280															
Db	2221	 aaacgcagc	gcgtgc	atcag	gttcac	atcgcagagctcaaaagcccaagcctacttt	2280															
Qy	2281	 caatgttac	ccgcacca	cttac	caat	tggtgtaaaacccgcagcttaacagtgatgg	2340															
Db	2281	 caatgttac	ccgcacca	cttac	caat	tggtgtaaaacccgcagcttaacagtgatgg	2340															
Qy	2341	 cactagtga	taaat	ttagt	gttta	agggttagtgtagcagacaataagcttagtaccgcga	2400															
Db	2341	 cactagtga	taaat	ttagt	gttta	agggttagtgtagcagacaataagcttagtaccgcga	2400															
Qy	2401	 acatttgg	caagctat	c	t	aaatggaagtc	aaatcg	acagctgacagtgctctcaaaagctt	2460													
Db	2401	 acatttgg	caagctat	c	t	aaatggaagtc	aaatcg	acagctgacagtgctctcaaaagctt	2460													
Qy	2461	 taccgttta	aaagaaga	gcagat	gtag	cccaacgctat	accgtggctggtc	aaagatacagac	2520													
Db	2461	 taccgttta	aaagaaga	gcagat	gtag	cccaacgctat	accgtggctggtc	aaagatacagac	2520													
Qy	2521	 aaaaaa	tgc	cggcag	ctcag	catctta	aaac	ctcaaaaggttaaaac	cggtctaac	2580												
Db	2521	 aaaaaa	tgc	cggcag	ctcag	catctta	aaac	ctcaaaaggttaaaac	cggtctaac	2580												
Qy	2581	 tacc	aaaaa	agat	tggt	ac	aggtt	ac	at	ggcgttag	cccaattgg	2640										
Db	2581	 tacc	aaaaa	agat	tggt	ac	aggtt	ac	at	ggcgttag	cccaattgg	2640										
Qy	2641	 caaaagc	ac	ccctaa	acac	gcag	tggt	gtta	agata	c	cccaac	caaa	2700									
Db	2641	 caaaagc	ac	ccctaa	acac	gcag	tggt	gtta	agata	c	cccaac	caaa	2700									
Qy	2701	 cgggtg	ctaa	tggc	att	t	acta	atgt	gaat	gttag	ta	tc	ccaggt	actgg	2760							
Db	2701	 cgggtg	ctaa	tggc	att	t	acta	atgt	gaat	gttag	ta	tc	ccaggt	actgg	2760							
Qy	2761	 aaat	t	ac	gcgt	ctgc	at	t	acc	agagata	aaaa	att	tg	gtt	ct	at	gg	tc	ag	tt	ga	2820
Db	2761	 aaat	t	ac	gcgt	ctgc	at	t	acc	agagata	aaaa	att	tg	gtt	ct	at	gg	tc	ag	tt	ga	2820
Qy	2821	 taca	aa	ca	aa	cc	ct	t	at	ct	g	at	ca	gc	a	a	gc	a	gc	a	a	2880
Db	2821	 taca	aa	ca	aa	cc	ct	t	at	ct	g	at	ca	gc	a	a	gc	a	gc	a	a	2880
Qy	2881	 cact	gg	ct	ta	t	aa	c	gc	g	g	gt	g	gt	g	gt	g	gt	g	gt	g	2940
Db	2881	 cact	gg	ct	ta	t	aa	c	gc	g	g	gt	g	gt	g	gt	g	gt	g	gt	g	2940

D	b	2881	cactggcattaacgcaggtggtgtaaaagccatcacagggtgtccccaacactgcctagcat	294
Q	y	2941	tgcgcataaagtagcgcgaacatagaactggcgcaatacaatccaagacaagaacaatc	3000
D	b	2941	tgcgcataaagtagcgcgaacatagaactggcgcaatacaatccaagacaagaacaatc	3000
Q	y	3001	caacgctgccagcattaatgatgatatataacacacagcgttttaacctataaaataatacaaa	3060
D	b	3001	caacgctgccagcattaatgatgatatataacacacagcgttttaacctataaaataatacaaa	3060
Q	y	3061	cccattgactttgtctccacttatgacattgttgactttgccaatggcaatgcaccac	3120
D	b	3061	cccattgactttgtctccacttatgacattgttgactttgccaatggcaatgcaccac	3120
Q	y	3121	cgccacagtcacccatgatcacgccttaacaaacacagtaaaagtggatatgatgtgaatgt	3180
D	b	3121	cgccacagtaacccatgatcacgccttaacaaacacagtaaaagtggatatgatgtgaatgt	3180
Q	y	3181	ggatgatacaaccattcatctaacagggcactgatgacataaaaaacttggcgtcaaaac	3240
D	b	3181	ggatgatacaaccattcatctaacagggcactgatgacataaaaaacttggcgtcaaaac	3240
Q	y	3241	cacaaactgaacaaacaaagtcgtaattggttaacagcaactaactttaatgttaactc	3300
D	b	3241	cacaaactgaacaaacaaagtcgtaattggttaacagcaactaactttaatgttaactc	3300
Q	y	3301	tagtgatgaagatgccttgttaacgccaaagacatgccgaaatctaaacacccctagc	3360
D	b	3301	tagtga tgaaga tgccttgttaacgccaaagacatgccgaaatctaaacacccctagc	3360
Q	y	3361	caaggaatttcacaccacaaagcgacagcacgcgcctcacaaactttaccgttaa	3420
D	b	3361	caaggaatttcacaccacaaagcgacagcacgcgcctcacaaactttaccgttaa	3420
Q	y	3421	aaaggtagatgaaataataatctgtatgacgccaaacgcatcccgtygggtcaaaagaa	3480
D	b	3421	aaaggtagatgaaataataatctgtatgacgccaaacgcatcccgtygggtcaaaagaa	3480
Q	y	3481	cgcaataatcaagtcaaacaccctaacctcaaaagtgaaacgctcttaatataaac	3540
D	b	3481	cgcaataatcaagtcaaacaccctaacctcaaaagtgaaacgctcttaatataaac	3540
Q	y	3541	cgacaaaattggtacggttacctttggcatttaacaccaacgcggtcttaagcgcgcaa	3600
D	b	3541	cgacaaaattggtacggttacctttggcatttaacaccaacgcggtcttaagcgcgcaa	3600
Q	y	3601	aagcacccctaaacgacggtggctgtctataaaaaacccactgtagcgacaataacca	3660
D	b	3601	aagcacccctaaacgacggtggctgtctataaaaaacccactgtagcgacaataacca	3660
Q	y	3661	agtcggttgctgatggcgtgaagtttgcgaagtttaataatgggtgtgttagtgtctgg	3720
D	b	3661	agtcggttgctgatggcgtgaagtttgcgaagtttaataatgggtgtgttagtgtctgg	3720
Q	y	3721	cattgatggcacactcgcatcaccagagatgaaattggcttactgggactaatggcttc	3780
D	b	3721	cattgatggcacactcgcatcaccagagatgaaattggcttactgggactaatggcttc	3780
Q	y	3781	acttgataaaagcaaaaccccaactaagcaaaagcagcattaacgcaggtggtaaaaagat	3840
D	b	3781	acttgataaaagcaaaaccccaactaagcaaaagcagcattaacgcaggtggtaaaaagat	3840
Q	y	3841	taccacattcaatcaggttgagattgccaaaacacagccaatgatgtgtgacagcgcgcaa	3900
D	b	3841	taccacattcaatcaggttgagattgccaaaacacagccaatgatgtgtgacagcgcgcaa	3900
Q	y	3901	gatttatgatttaaaacacgaaacttgaaacaaaaatcagcagctactgccaaaacagcaca	3960
D	b	3901	gatttatgatttaaaacacgaaacttgaaacaaaaatcagcagctactgccaaaacagcaca	3960
Q	y	3961	aaactcattacagaaattctcagtagcagatgaacaaaggtataaactttacggttagtaa	4020
D	b	3961	aaactcattacagaaattctcagtagcagatgaacaaaggtataaactttacggttagtaa	4020


```

Db 6181 cgcaggcacacagcaagcaaaaaatctgacggcacagcagggtacaaaccaccacagcagggtgc 6240
Qy 6241 aacgggtacggttaaaagcgttctgtgacaaacggcggttgggtgcgtctccgtgggtgc 6300
Db 6241 aacgggtacggttaaaagcgttctgtgacaaacggcggttgggtgcgtctccgtgggtgc 6300
Qy 6301 ctcaagggtcgaacgcgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagcac 6360
Db 6301 ctcaagggtcgaacgcgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagcac 6360
Qy 6361 cgatgcggtcaatggtagcaggtgtacaaagccaccacaaagcatttgcacacgcaaccaa 6420
Db 6361 cgatgcggtcaatggtagcaggtgtacaaagccaccacaaagcatttgcacacgcaaccaa 6420
Qy 6421 tgagcttgaccatcgtatccaccacaaacgaaaaataagggccaaatgcagggtttcatcagc 6480
Db 6421 tgagcttgaccatcgtatccaccacaaacgaaaaataagggccaaatgcagggtttcatcagc 6480
Qy 6481 gatggcgatgggtccatgcacaaagcctacattcctctggcagatccatggttaccggggg 6540
Db 6481 gatggcgatgggtccatgcacaaagcctacattcctctggcagatccatggttaccggggg 6540
Qy 6541 tattgccaccacaaacggtcaaggtgcggtggcagtgaggactgtcgaagctgtcgggataa 6600
Db 6541 tattgccaccacaaacggtcaaggtgcggtggcagtgaggactgtcgaagctgtcgggataa 6600
Qy 6601 tgggtcaatgggttataaaatcaatggttcagcgcgatacccaaggccatgtagggcggc 6660
Db 6601 tgggtcaatgggttataaaatcaatggttcagcgcgatacccaaggccatgtagggcggc 6660
Qy 6661 agttgggtcaggttttcaacttttaagcgcataaaatcgaagattttacttataaaatcaatc 6720
Db 6661 agttgggtcaggttttcaacttttaagcgcataaaatcgaagattttacttataaaatcaatc 6720
Qy 6721 tcaccatagttgtataaaacagcatcagcatcagtcataattactgagtggtgttttt 6780
Db 6721 tcaccatagttgtataaaacagcatcagcatcagtcataattactgagtggtgttttt 6780
Qy 6781 atcacttaaacattttaccgctcaagtgattctcttccaccatgaccaaatgcgaattg 6840
Db 6781 atcacttaaacattttaccgctcaagtgattctcttccaccatgaccaaatgcgaattg 6840
Qy 6841 atcatagtgaaacttctgagtaaaatttatcaatgtagttagttgatatggttaaaatt 6900
Db 6841 atcatagtgaaacttctgagtaaaatttatcaatgtagttagttgatatggttaaaatt 6900
Qy 6901 gtgcattgaccaaataatgacgattttatcccgaaatttctgattatgattccgttgac 6960
Db 6901 gtgcattgaccaaataatgacgattttatcccgaaatttctgattatgattccgttgac 6960
Qy 6961 ctgcaggtcgac 6972
Db 6961 ctgcaggtcgac 6972

```

RESULT 2

```

AAF59100
ID AAF59100 standard; DNA; 6973 BP.
XX
AC
AC AAF59100;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene SEQ ID NO:1.
XX
KW Moraxella catarrhalis strain 4223; major outer membrane protein;
XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW oclitis media; detection; ds.
XX
OS Moraxella catarrhalis.
XX
XX W0200107619-A1.
PN

```

```

XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loomore SM, Sasaki K, Yang Y, Klein MH;
XX
WPI: 2001-159722/16.
XX
P-PSDB; AAB59133.
XX
PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
XX Example 3; Fig 2A-W; 247pp; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene, which is used
XX in the exemplification of the present invention.
XX
SQ Sequence 6973 BP; 2265 A; 1555 C; 1533 G; 1620 T; 0 other;

```

Query Match 99.8%; Score 6961; DB 22; Length 6973;

Best Local Similarity 100.0%; Pred. NO. 0;

Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 ccattgatatggcaggtgtgctgcctgcctgattgagcgatgacacccattgtgcc 60

Db 1 ccattgatatggcaggtgtgctgcctgcctgattgagcgatgacacccattgtgcc 60

Qy 61 catactgtacgatttgacatgtgatatttaacatgtgacatgatttaacattgttt 120

Db 61 catactgtacgatttgacatgtgatatttaacatgtgacatgatttaacattgttt 120

Qy 121 aatactgttgcattatcaacataatttagtaacgcgatttagtaacgcaattgttaaaat 180

Db 121 aatactgttgcattatcaacataatttagtaacgcgatttagtaacgcaattgttaaaat 180

Qy 181 cattgcgccccctttatgtatcatatgaatagaatattatgattgtatcgtatttgt 240

Db 181 cattgcgccccctttatgtatcatatgaatagaatattatgattgtatcgtatttgt 240

Qy 241 atcagaatggtgtgattgatgtgcctacgagtgattggtggttaacactctatg 300

Db 241 atcagaatggtgtgattgatgtgcctacgagtgattggtggttaacactctatg 300

Qy 301 attgatatattttgaaactaaatctattgacttaaacaccatattggtataatttagca 360

Db 301 attgatatattttgaaactaaatctattgacttaaacaccatattggtataatttagca 360

Qy 361 taatggtgaggtttttgtaaaataacacatgcgaatattgttctactgttactacatgct 420

Db 361 taatggtgaggtttttgtaaaataacacatgcgaatattgttctactgttactacatgct 420

Qy 421 tgaatgacgatcccaatcaccagatttcattcaagtgtgtgtttgttacgcaccattta 480

Db 421 tgaatgacgatcccaatcaccagatttcattcaagtgtgtgtttgttacgcaccattta 480

QY 481 cccatattatttcaatcaaatgccttatgtcagcatgtatcatcttttttaaggtaaacac 540
 Db 481 cccatattatttcaatcaaatgccttatgtcagcatgtatcatcttttttaaggtaaacac 540
 QY 541 catgaatcacatctataaagtcatcttttaacaaagccacagccacatttatggcagtgcc 600
 Db 541 catgaatcacatctataaagtcatcttttaacaaagccacagccacatttatggcagtgcc 600
 QY 601 agagtacgccaatccacagcac-gggggggggtagctgtgtctacagggccaagtggca 659
 Db 601 agagtacgccaatccacagcacgggggggggtagctgtgtctacagggccaagtggca 660
 QY 660 gtgtatgcactctgagcttgcogtattgcccgcctgcgtgcgtctcctgtatcggtgcaa 719
 Db 661 gtgtatgcactctgagcttgcogtattgcccgcctgcgtgcgtctcctgtatcggtgcaa 720
 QY 720 cgtcagtgagcagtgcttatgtctcaaaaaaagataccaaacatatcgcaattggtgaac 779
 Db 721 cgtcagtgagcagtgcttatgtctcaaaaaaagataccaaacatatcgcaattggtgaac 780
 QY 780 aaacacagccaagacgctcaggaactgcgaagccgaagccggtgatcgagccattgctattg 839
 Db 781 aaacacagccaagacgctcaggaactgcgaagccgaagccggtgatcgagccattgctattg 840
 QY 840 gtgaaaaatgctaaccgacagggcggtcaagccatcgccatcggttagtagtaataaaactg 899
 Db 841 gtgaaaaatgctaaccgacagggcggtcaagccatcgccatcggttagtagtaataaaactg 900
 QY 900 tcaatggaagcagtttggataagtagtaggtaccgatgtctacgggtcaagagtcacatgcga 959
 Db 901 tcaatggaagcagtttggataagtagtaggtaccgatgtctacgggtcaagagtcacatgcga 960
 QY 960 tcgggtggtagttaaaggctagtggtgatgcctcgatgccatcggttagtagtgcattac 1019
 Db 961 tcgggtggtagttaaaggctagtggtgatgcctcgatgccatcggttagtagtgcattac 1020
 QY 1020 atttgcctgtagcagtggttaactcctaaacatccgaaggtactctgattaaagcttta 1079
 Db 1021 atttgcctgtagcagtggttaactcctaaacatccgaaggtactctgattaaagcttta 1080
 QY 1080 ttaacggccatgcagttataaagaaatacgaagctcaaaagataatgatgtaaaatata 1139
 Db 1081 ttaacggccatgcagttataaagaaatacgaagctcaaaagataatgatgtaaaatata 1140
 QY 1140 gacgcacaacgcgaagcgacagccagtagtactgcagtgggagccatgtcatatgcacag 1199
 Db 1141 gacgcacaacgcgaagcgacagccagtagtactgcagtgggagccatgtcatatgcacag 1200
 QY 1200 gtcatattttccaaacycctttggtacacgggcaacagctaaaagtgcctattccttggcag 1259
 Db 1201 gtcatattttccaaacycctttggtacacgggcaacagctaaaagtgcctattccttggcag 1260
 QY 1260 tgggtcttgcgcacagcgagggccaactcaaatcgctattggttctgtatgcaacat 1319
 Db 1261 tgggtcttgcgcacagcgagggccaactcaaatcgctattggttctgtatgcaacat 1320
 QY 1320 ctatgcctgttgggagcagatagcccttgggtgcaggtactcgtctcagctacagggcagta 1379
 Db 1321 ctatgcctgttgggagcagatagcccttgggtgcaggtactcgtctcagctacagggcagta 1380
 QY 1380 ttgcccctaggtcaagggttctgttctactcagagtgataataattctagaccggcctata 1439
 Db 1381 ttgcccctaggtcaagggttctgttctactcagagtgataataattctagaccggcctata 1440
 QY 1440 caccataatcccagggcactagaccccaagtttcaagccaccaataataacgaagcggggtc 1499
 Db 1441 caccataatcccagggcactagaccccaagtttcaagccaccaataataacgaagcggggtc 1500
 QY 1500 cactttccattggttagtaactctatcaaaacgtaaaatacatcaatgctcgggtcaggtgta 1559
 Db 1501 cactttccattggttagtaactctatcaaaacgtaaaatacatcaatgctcgggtcaggtgta 1560

QY 1560 ataaacccgagtcggtcacaatgtggcagcagctagaagcgtgtggaagtggtgctaaggagc 1619
 Db 1561 ataaacccgagtcggtcacaatgtggcagcagctagaagcgtgtggaagtggtgctaaggagc 1620
 QY 1620 gtagaattacttttcaggggtgatgataaacagtagtactgacgtataaaatagggttttggaata 1679
 Db 1621 gtagaattacttttcaggggtgatgataaacagtagtactgacgtataaaatagggttttggaata 1680
 QY 1680 ctttaactattaaagggtgtgcagagaccacgcattaaaccgataataatcgttgcgtgtg 1739
 Db 1681 ctttaactattaaagggtgtgcagagaccacgcattaaaccgataataatcgttgcgtgtg 1740
 QY 1740 taaaagggtcgtataatagtggtctgaaagttaaactgtctaaactgtctaaactttaacaacttta 1799
 Db 1741 taaaagggtcgtataatagtggtctgaaagttaaactgtctaaactgtctaaactttaacaacttta 1800
 QY 1800 ctgagggtgaatacacactacattaaatgcccaacacacagttaaggttaggttagtagtagta 1859
 Db 1801 ctgagggtgaatacacactacattaaatgcccaacacacagttaaggttaggttagtagtagta 1860
 QY 1860 gtactacagctgaattatttgagtgatagtttaacctttaccagccccaataacagggcagtc 1919
 Db 1861 gtactacagctgaattatttgagtgatagtttaacctttaccagccccaataacagggcagtc 1920
 QY 1920 aaagcacaagcaaaacccgtctatggcgttaattgggtggaagttaacttaataatgcagaaa 1979
 Db 1921 aaagcacaagcaaaacccgtctatggcgttaattgggtggaagttaacttaataatgcagaaa 1980
 QY 1980 caacagcagcaatcgccactactcgtattaccagagataaaatggcgtttgtctcgcagatg 2039
 Db 1981 caacagcagcaatcgccactactcgtattaccagagataaaatggcgtttgtctcgcagatg 2040
 QY 2040 gtgatgtgtgataaaaaacaagcaccatatttggataaaaaacaacttaaaatgggttagtg 2099
 Db 2041 gtgatgtgtgataaaaaacaagcaccatatttggataaaaaacaacttaaaatgggttagtg 2100
 QY 2100 ttgcaattaccatagacagatgctgctgaggttaataaaaaagatcagtaactcttgcca 2159
 Db 2101 ttgcaattaccatagacagatgctgctgaggttaataaaaaagatcagtaactcttgcca 2160
 QY 2160 aaggtagcagtgctaacgctggtttaccatcgaaacagctcaaaagccgccaagcctaactt 2219
 Db 2161 aaggtagcagtgctaacgctggtttaccatcgaaacagctcaaaagccgccaagcctaactt 2220
 QY 2220 taaaagcagcgctgcagctacgtgtcaacccctactgaaataatcagttgatgctaagagtg 2279
 Db 2221 taaaagcagcgctgcagctacgtgtcaacccctactgaaataatcagttgatgctaagagtg 2280
 QY 2280 gcaatgttaccgcccacaacttacaacattggcgtgaaaccccgagcttaacagtgatg 2339
 Db 2281 gcaatgttaccgcccacaacttacaacattggcgtgaaaccccgagcttaacagtgatg 2340
 QY 2340 gcaatgtatgaataatttagtgttaagggtagtggtacgacaataatagcttagttaccgccc 2399
 Db 2341 gcaatgtatgaataatttagtgttaagggtagtggtacgacaataatagcttagttaccgccc 2400
 QY 2400 aacatttggcaagctatctaaatgaagtcaatcgaaacggtgacagtgctctacaaagct 2459
 Db 2401 aacatttggcaagctatctaaatgaagtcaatcgaaacggtgacagtgctctacaaagct 2460
 QY 2460 ttaccgtttaaagaagaagcagatgtagcgccaacgctatcccggtggtgctaagatacga 2519
 Db 2461 ttaccgtttaaagaagaagcagatgtagcgccaacgctatcccggtggtgctaagatacga 2520
 QY 2520 caaaaaatcccgccagtcagctacatcttaaaactcaaaaggttaaaacggtctctaaacggttg 2579
 Db 2521 caaaaaatcccgccagtcagctacatcttaaaactcaaaaggttaaaacggtctctaaacggttg 2580
 QY 2580 ctaccaaaaaagatggtacggtttacctttggggttagccaagatcagcgtctgacacatg 2639
 Db 2581 ctaccaaaaaagatggtacggtttacctttggggttagccaagatcagcgtctgacacatg 2640
 QY 2640 gcaaaagcacccttaaaacacgagtggttgaagtataccaaacgaacaaatccaaag 2699

Db 2641 gcaaaagcacccataaacaacgagtggtgactggttaagatacacaacgaacaaatccaag 2700
Qy 2700 tcggtgctaattgacattactaaattactaatgtgaatggttagtaataccaggtactggcattg 2759
Db 2701 tcggtgctaattgacattactaaattactaatgtgaatggttagtaataccaggtactggcattg 2760
Qy 2760 caaataccgctgcgattaccagagataaaaattggctttgctgtctgtatggttgagcttg 2819
Db 2761 caaataccgctgcgattaccagagataaaaattggctttgctgtctgtatggttgagcttg 2820
Qy 2820 atacaaacaacatttacttgatcaagaacagctacaagttggcaatgttaagattacca 2879
Db 2821 atacaaacaacatttacttgatcaagaacagctacaagttggcaatgttaagattacca 2880
Qy 2880 aactggtcattaaacgaggtggttaaacccatcacaggtgtgcccaacactgacctagca 2939
Db 2881 aactggtcattaaacgaggtggttaaacccatcacaggtgtgcccaacactgacctagca 2940
Qy 2940 ttgcgatacaagtagccgcaacatagaactgggcaatacaataccaaagacaacaaat 2999
Db 2941 ttgcgatacaagtagccgcaacatagaactgggcaatacaataccaaagacaacaaat 3000
Qy 3000 ccaacgctgcagcattaatgatattataataacaggtttaaccttaaaaaataataaca 3059
Db 3001 ccaacgctgcagcattaatgatattataataacaggtttaaccttaaaaaataataaca 3060
Qy 3060 acccattgactttgtctccacttatgacattgttgactttgccaatggcaatggccacca 3119
Db 3061 acccattgactttgtctccacttatgacattgttgactttgccaatggcaatggccacca 3120
Qy 3120 ccgcccagtaacccatgataccogctaacaaacagtaaaagtggtatagtgtgaatg 3179
Db 3121 ccgcccagtaacccatgataccogctaacaaacagtaaaagtggtatagtgtgaatg 3180
Qy 3180 tggatgatacaaccattcatctaacaggaactgatgacaataaaaaacttggcgtaaaaa 3239
Db 3181 tggatgatacaaccattcatctaacagggactgatgacaataaaaaacttggcgtaaaaa 3240
Qy 3240 ccaccaactgaacaaacaaagtgctaattggttaatacagcaactaaactttaattgtaact 3299
Db 3241 ccaccaactgaacaaacaaagtgctaattggttaatacagcaactaaactttaattgtaact 3300
Qy 3300 ctagtgtgaagtgccttgtttaacgccaaagacatacgccgaaaaatctaaacacccctag 3359
Db 3301 ctagtgtgaagtgccttgtttaacgccaaagacatacgccgaaaaatctaaacacccctag 3360
Qy 3360 ccaaggaaattcacaccacaaaggacagacagaccccgccctacaaacottttaccggtta 3419
Db 3361 ccaaggaaattcacaccacaaaggacagacaccccgccctacaaacottttaccggtta 3420
Qy 3420 aaaaggtagatgaaaaataataatgctgatgacgccaacgccaatccacggtgggtcaaaaga 3479
Db 3421 aaaaggtagatgaaaaataataatgctgatgacgccaacgccaatccacggtgggtcaaaaga 3480
Qy 3480 acgcaataatcaagtcaacacccctaaactcaaaagtgaacacggttcttaattataaaa 3539
Db 3481 acgcaataatcaagtcaacacccctaaactcaaaagtgaacacggttcttaattataaaa 3540
Qy 3540 ccgaacaaattggttacggtttacctttggcatttaacacccaagcggctttaaagcccgca 3599
Db 3541 ccgaacaaattggttacggtttacctttggcatttaacacccaagcggctttaaagcccgca 3600
Qy 3600 aaagcacccctaaaacgacggtggcttgcattataaaaaaccccaactggtagcggaacaaatcc 3659
Db 3601 aaagcacccctaaaacgacggtggcttgcattataaaaaaccccaactggtagcggaacaaatcc 3660
Qy 3660 aagtcggtgctgataggcgtgaagtgttccaaaggttaataataatggttgttaggtgctg 3719
Db 3661 aagtcggtgctgataggcgtgaagtgttccaaaggttaataataatggttgttaggtgctg 3720
Qy 3720 gcatgtatggcacaaactcgcattaccagagataaaattggctttacttgggactaatggct 3779

Db 3721 gcatgtatggcacaaactcgcattaccagagatgaattggctttacttgggactaatggct 3780
Qy 3780 cacttgataaaagcaaacccccacacccacctaagcaaacgacgcttaacacgaggttggttaaaaga 3839
Db 3781 cacttgataaaagcaaacccccacacccacctaagcaaacgacgcttaacacgaggttggttaaaaga 3840
Qy 3840 ttaccaacattcaatcagggtgagattgtcccaaaacacgccatgactgtgtgacagggcgca 3899
Db 3841 ttaccaacattcaatcagggtgagattgtcccaaaacacgccatgactgtgtgacagggcgca 3900
Qy 3900 agatttatgatttaaaacccgaacttgaacaaaatacagcagctactgcgaacacagcac 3959
Db 3901 agatttatgatttaaaacccgaacttgaacaaaatacagcagctactgcgaacacagcac 3960
Qy 3960 aaaactcattacacgaattctcagtagcagatgaacaaaggttaataacttttacggttagta 4019
Db 3961 aaaactcattacacgaattctcagtagcagatgaacaaaggttaataacttttacggttagta 4020
Qy 4020 acccttactcagttatgacacctcgaagacctctgatgtcatcacctttgcaggtgaaa 4079
Db 4021 acccttactcagttatgacacctcgaagacctctgatgtcatcacctttgcaggtgaaa 4080
Qy 4080 acggcattaccacacgaaggttaataaaggtgtgtgctgtgtggcattgacaaacacaaag 4139
Db 4081 acggcattaccacacgaaggttaataaaggtgtgtgctgtgtggcattgacaaacacaaag 4140
Qy 4140 gcttaacacgcttaagctgacggtggttaataataatggcaagggcattgtcattgaca 4199
Db 4141 gcttaacacgcttaagctgacggtggttaataataatggcaagggcattgtcattgaca 4200
Qy 4200 gccaaaatggttcaaaaataccatcacaggaacttaagcaacactctagctaatgttaccacatg 4259
Db 4201 gccaaaatggttcaaaaataccatcacaggaacttaagcaacactctagctaatgttaccacatg 4260
Qy 4260 ataaagtagcgttacgcaccacagacagggccaataatacaaaagagaagacaaaacccc 4319
Db 4261 ataaagtagcgttacgcaccacagacagggccaataatacaaaagagaagacaaaacccc 4320
Qy 4320 gtgcgcgcagcattgtgtgctaaagcgcagcgtttaaacttgaaggaacgaatggtgaag 4379
Db 4321 gtgcgcgcagcattgtgtgctaaagcgcagcgtttaaacttgaaggaacgaatggtgaag 4380
Qy 4380 cggttgactttgtctccacttatgacacccgttcaactttgcgatggcaatgccaccacgg 4439
Db 4381 cggttgactttgtctccacttatgacacccgttcaactttgcgatggcaatgccaccacgg 4440
Qy 4440 ctaaggtgacctatgatacacaaagcaaacacagtaaaagtggtctatgatactgaatgtgg 4499
Db 4441 ctaaggtgacctatgatacacaaagcaaacacagtaaaagtggtctatgatactgaatgtgg 4500
Qy 4500 atgatacaacccattgaagttaaagataaaaaacttggcgttaaaaaacccacacattgacca 4559
Db 4501 atgatacaacccattgaagttaaagataaaaaacttggcgttaaaaaacccacacattgacca 4560
Qy 4560 gtactggcacaggttgcttaataaaatttgccttaagcaatcaagctacttggcgatgcgcttg 4619
Db 4561 gtactggcacaggttgcttaataaaatttgccttaagcaatcaagctacttggcgatgcgcttg 4620
Qy 4620 tcaagggcaggtgactgtgctcatcaaaccttatctggcgacatccaaactgcca 4679
Db 4621 tcaagggcaggtgactgtgctcatcaaaccttatctggcgacatccaaactgcca 4680
Qy 4680 aaggggcaagccaagccaacaaactcagcaggtctgtggaatggcgaataaaggttca 4739
Db 4681 aaggggcaagccaagccaacaaactcagcaggtctgtggaatggcgaataaaggttca 4740
Qy 4740 tctatgacagtaccgatacaaaagtactatcaagccaaaaatgataagcaggttgataaaa 4799
Db 4741 tctatgacagtaccgatacaaaagtactatcaagccaaaaatgataagcaggttgataaaa 4800
Qy 4800 ccaaaagaggttgcacaaagcaaaactggtcgcccaagcccaaacccagatggcacattgg 4859
Db 4801 ccaaaagaggttgcacaaagcaaaactggtcgcccaagcccaaacccagatggcacattgg 4860

Qy	4860	ctcaaatgaatgtcaaatcagtcatttaacaaagaaacaagttaaatgatgcgaataaaaagc	4919
Db	4861	ctcaaatgaatgtcaaatcagtcatttaacaaagaaacaagttaaatgatgcgaataaaaagc	4920
	4920	aaggcatcaatgaagacaacacgcctttgttaaaggacttgaaaagcgcgtctctgataaca	4979
Db	4921	aaggcatcaatgaagacaacacgcctttgttaaaggacttgaaaagcgcgtctctgataaca	4980
Qy	4980	aaacaaaaacgcgcgagtaactgtgggtgatttaaatgcgcttgcccaaacacccgtga	5039
Db	4981	aaacaaaaacgcgcgagtaactgtgggtgatttaaatgcgcttgcccaaacacccgtga	5040
Qy	5040	cccttgacgggatacagggcacacgcgctaaaaaacctggcgagagacttgcacatcaaa	5099
Db	5041	cccttgacgggatacagggcacacgcgctaaaaaacctggcgagagacttgcacatcaaa	5100
Qy	5100	gtgggcacaaacagacacaaataagctaaaccgataataacatcggctgtgtgtagcaggtactg	5159
Db	5101	gtgggcacaaacagacacaaataagctaaaccgataataacatcggctgtgtgtagcaggtactg	5160
Qy	5160	atggtcttcactgtccaaacttgccaaagacctaaccgaatctttaacagcgcttaaatcaggtg	5219
Db	5161	atggtcttcactgtccaaacttgccaaagacctaaccgaatctttaacagcgcttaaatcaggtg	5220
Qy	5220	gcacaaataatgatgacaaagcgtgtctttttagactcaagcggctcaagcccaaacgcaa	5279
Db	5221	gcacaaataatgatgacaaagcgtgtctttttagactcaagcggctcaagcccaaacgcaa	5280
Qy	5280	acacccctgtctaaagtgcgaatggcctggacctgggtggcgaagtcacagtaaatgtcg	5339
Db	5281	acacccctgtctaaagtgcgaatggcctggacctgggtggcgaagtcacagtaaatgtcg	5340
Qy	5340	gcaagggcacaaaagataccgacgctgccaatgtacacagcttaaacgaagttaacgcaact	5399
Db	5341	gcaagggcacaaaagataccgacgctgccaatgtacacagcttaaacgaagttaacgcaact	5400
Qy	5400	tgttggtcttgttaatgctggtaatgataacgctgaacggaatcaggtataacatgtccg	5459
Db	5401	tgttggtcttgttaatgctggtaatgataacgctgaacggaatcaggtataacatgtccg	5460
Qy	5460	acatacaaaaaagacccaaatcaggttcaatcctaaccgcactgcatacaaacgagcca	5519
Db	5461	acatacaaaaaagacccaaatcaggttcaatcctaaccgcactgcatacaaacgagcca	5520
Qy	5520	cggctacttgccgttaaggtataacgataccgcaaaaacttgccactggtggtatacaag	5579
Db	5521	cggctacttgccgttaaggtataacgataccgcaaaaacttgccactggtggtatacaag	5580
Qy	5580	tgggcgtggtataaagacgcgcacgcgttaacgcgatttaagcaaatgtttgggttcaaaccc	5639
Db	5581	tgggcgtggtataaagacgcgcacgcgttaacgcgatttaagcaaatgtttgggttcaaaccc	5640
Qy	5640	aaaaaatggcaacaaaaagccctgctgcacacttataacgcgcgaggtcgacaccaact	5699
Db	5641	aaaaaatggcaacaaaaagccctgctgcacacttataacgcgcgaggtcgacaccaact	5700
Qy	5700	atttgaccaaaccccgagagccattgacagaataaatgaacaggtatccgcttct	5759
Db	5701	atttgaccaaaccccgagagccattgacagaataaatgaacaggtatccgcttct	5760
Qy	5760	tcacatgcaacgatggcaatcaagagccctgtgtacaaagggcgttaacgcgcatctgactcaa	5819
Db	5761	tcacatgcaacgatggcaatcaagagccctgtgtgtacaaagggcgttaacgcgcatctgactcaa	5820
Qy	5820	gtgcctcaggcaaacgactcagttggcgatagtttccaggccaagggcagatgggtgaagccg	5879
Db	5821	gtgcctcaggcaaacgactcagttggcgatagtttccaggccaagggcagatgggtgaagccg	5880
Qy	5880	ccgttgccataggcgagacaaccccaagcagggcaaccaatccatcgcccatcgggtgataacg	5939
Db	5881	ccgttgccataggcgagacaaccccaagcagggcaaccaatccatcgcccatcgggtgataacg	5940

Qy	5940	cacaagccacggcgagatcaatccactcgccatcggtacaggcaatgtggtacgaggtaaagc	5999
Db	5941	cacaagccacggcgagatcaatccactcgccatcggtacaggcaatgtggtacgaggtaaagc	6000
Qy	6000	actctggtgcacatcggcgagaccgaagcactgttaagctgataacagttacagttggtggta	6059
Db	6001	actctggtgcacatcggcgagaccgaagcactgttaagctgataacagttacagttggtggta	6060
Qy	6060	ataacaaaccagtttaccgatgcactcaaacccgactgtcttgggtgggcaataaataca	6119
Db	6061	ataacaaaccagtttaccgatgcactcaaacccgactgtcttgggtgggcaataaataca	6120
Qy	6120	cogtgaccggaagaataactcgtgttccttaggttcaaacctctgcacatcagtgacagcacac	6179
Db	6121	cogtgaccggaagaataactcgtgttccttaggttcaaacctctgcacatcagtgacagcacac	6180
Qy	6180	acgcagggcacacaagccaaaatactgcggcagcagcggtacaaacacacacagcaggtg	6239
Db	6181	acgcagggcacacaagccaaaatactgcggcagcagcggtacaaacacacacagcaggtg	6240
Qy	6240	caaccggtacggttaaaagccttgctggacaaaacggcggttgggtcggtctccgtgggtg	6299
Db	6241	caaccggtacggttaaaagccttgctggacaaaacggcggttgggtcggtctccgtgggtg	6300
Qy	6300	cctcaggtgctgaacgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagca	6359
Db	6301	cctcaggtgctgaacgcgtatccaaaatgtggcagcaggtgaggtcagtgccaccagca	6360
Qy	6360	cggatcggtcgaatggtagccagttgtacaagaagccaccaaagcattgccaaacgcaacca	6419
Db	6361	cggatcggtcgaatggtagccagttgtacaagaagccaccaaagcattgccaaacgcaacca	6420
Qy	6420	atgagcttgacacatcgtatccaccaaaacgaataaagcccaatgcagggatttcatacag	6479
Db	6421	atgagcttgacacatcgtatccaccaaaacgaataaagcccaatgcagggatttcatacag	6480
Qy	6480	cgatggcgatggcgctccatgcacaaagcctacattcctggcgagatccatggttaccgggg	6539
Db	6481	cgatggcgatggcgctccatgcacaaagcctacattcctggcgagatccatggttaccgggg	6540
Qy	6540	gtattgccaccacacacggtcaaggctcggtggcagtggaagctgcgaagctgcggata	6599
Db	6541	gtattgccaccacacacggtcaaggctcggtggcagtggaagctgcgaagctgcggata	6600
Qy	6600	atggccaatgggtatttaaaatcaatggttcagccgatacccaaagccatgtaggggcgg	6659
Db	6601	atggccaatgggtatttaaaatcaatggttcagccgatacccaaagccatgtaggggcgg	6660
Qy	6660	cagttgggtcaggttttcaacttttaagccataaatgcgaagatttcacttaaaaaatcaat	6719
Db	6661	cagttgggtcaggttttcaacttttaagccataaatgcgaagatttcacttaaaaaatcaat	6720
Qy	6720	ctcacatagttgtataaaacagcatcagcatcagtcataattactgatgctgatgtttt	6779
Db	6721	ctcacatagttgtataaaacagcatcagcatcagtcataattactgatgctgatgtttt	6780
Qy	6780	tataccttaaacatttttacgcgtccaagtgattctcttccaccatgacaaaatcgccatt	6839
Db	6781	tataccttaaacatttttacgcgtccaagtgattctcttccaccatgacaaaatcgccatt	6840
Qy	6840	gatacaggttaaaccttattgagtaaaatttatcaatgattgttagatatggttaaaat	6899
Db	6841	gatacaggttaaaccttattgagtaaaatttatcaatgattgttagatatggttaaaat	6900
Qy	6900	tgtgccattgacaaaaaaatgacggatttatcccgaaaaattcttgattatgataccgttga	6959
Db	6901	tgtgccattgacaaaaaaatgacggatttatcccgaaaaattcttgattatgataccgttga	6960
Qy	6960	cctgcaggtcgac	6972
Db	6961	cctgcaggtcgac	6973

RESULT 3

```
AAT38740
ID   AAT38740 standard; DNA; 6971 BP.
XX   AC
XX   AAT38740;
XX   DT
XX   25-JAN-1997 (first entry)
XX   DE
XX   Moraxella outer membrane protein gene.
XX   KW
XX   Outer membrane protein; OMP; immunogen; vaccine; otitis media;
XX   diagnosis; ss.
XX   OS
XX   Mycobacterium catarrhalis strain 4223.
XX   FH
XX   Key
XX   Location/Qualifiers
XX   CDS
XX   706..6684
XX   /*tag= a
XX   PN
XX   W09634960-A1.
XX   PD
XX   07-NOV-1996.
XX   PF
XX   29-APR-1996; 96WO-CA00264.
XX   PR
XX   26-MAR-1996; 96US-0621944.
XX   PR
XX   01-MAY-1995; 95US-0431718.
XX   PR
XX   07-JUN-1995; 95US-0478370.
XX   PA
XX   (CONN-) CONNAUGHT LAB LTD.
XX   PI
XX   Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX   WPI; 1996-506162/50.
XX   DR
XX   P-PSDB; AAW04505.
XX   PT
XX   Moraxella outer membrane protein - useful as immunogen in protective
XX   vaccine and for diagnosis
XX   PS
XX   Claim 14; Fig 6; 109pp; English.
XX   CC
XX   A genomic DNA fragment (738740) of Moraxella catarrhalis otitis
XX   media strain 4223 includes the coding region for a 1992-amino
XX   acid protein (W04505) identified as an approx. 200 kDa outer
XX   membrane protein (OMP). The DNA was isolated from a strain 4223
XX   genomic library in phage lambda EMBL3 by screening with an
XX   anti-200 kDa protein guinea pig antiserum. The gene can be
XX   used for the recombinant expression of the OMP (for use in
XX   vaccines), for the prepn. of hybridisation probes, or may be
XX   incorporated into a live vector for use in direct immunisation.
XX   SQ
XX   Sequence 6971 BP; 2265 A; 1553 C; 1533 G; 1620 T; 0 other;

Query Match          99.88; Score 6959; DB 17; Length 6971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6970; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY   3 atggtatggcagggtgtcgcctgcctgtatgtagcgcatgacacccattggcccca 62
DB   1 atggtatggcagggtgtcgcctgcctgtatgtagcgcatgacacccattggcccca 60

QY   63 tatctgtacgatttgacattgacattgatttaacattgacattgatttaacattgttaa 122
DB   61 tatctgtacgatttgacattgacattgatttaacattgacattgatttaacattgttaa 120

QY   123 tactgttgccatcattaccataattttagtaacgcattagtaacgcatttgtaaaaaatca 182
DB   121 tactgttgccatcattaccataattttagtaacgcattagtaacgcatttgtaaaaaatca 180

QY   183 ttgcgccccctttatgtatcatatgaatagaattattatgtatctgattatttat 242
DB   181 ttgcgccccctttatgtatcatatgaatagaattattatgtatctgattatttat 240
```

```
QY   243 cagaatggtgatgctatgatgctacgagttgatttgggttaactactctatgat 302
DB   241 cagaatggtgatgctatgatgctacgagttgatttgggttaactactctatgat 300

QY   303 ttgatatttttgaactaatctattgacttaaatcaaccatatggttataatttagcata 362
DB   301 ttgatatttttgaactaatctattgacttaaatcaaccatatggttataatttagcata 360

QY   363 atggtaggctttttgtaaaaatcacatcgcaatttgttctactgttactaccatgcttg 422
DB   361 atggtaggctttttgtaaaaatcacatcgcaatttgttctactgttactaccatgcttg 420

QY   423 aatgacgatcccaatcaccagattcattcaagtgtgtttgttatcgaccattttacc 482
DB   421 aatgacgatcccaatcaccagattcattcaagtgtgtttgttatcgaccattttacc 480

QY   483 ctaattatttcaaatcaaatgctatgtcagcatgtatcattttttaaaggtaaacaccca 542
DB   481 ctaattatttcaaatcaaatgctatgtcagcatgtatcattttttaaaggtaaacaccca 540

QY   543 tgaatcacatctataaagtcatctttaaaaagccacagccacatttatggcagtgccag 602
DB   541 tgaatcacatctataaagtcatctttaaaaagccacagccacatttatggcagtgccag 600

QY   603 agtacgcaaatccacagcacac-ggggggggggtagctgtgtacagggcaagttggcagt 661
DB   601 agtacgcaaatccacagcacacgggggggggtagctgtgtacagggcaagttggcagt 660

QY   662 gtatgcactctgagctttgcccgatttgcgcgtgtgtcctcgtgtagcggtagcaacg 721
DB   661 gtatgcactctgagctttgcccgatttgcgcgtgtgtcctcgtgtagcggtagcaacg 720

QY   722 ctcaagtggcagtgcttatgtctcaaaaaaagatcaccaaacatcgcgaattggtgaacaa 781
DB   721 ctcaagtggcagtgcttatgtctcaaaaaaagatcaccaaacatcgcgaattggtgaacaa 780

QY   782 aaccagccaagcagctcaggcactgccaaggcgacggtagcggtagcggcatttgcattggt 841
DB   781 aaccagccaagcagctcaggcactgccaaggcgacggtagcggtagcggcatttgcattggt 840

QY   842 gaaatgtaacgacacagggcggtcaagccatcgccatcgtagtagtaataaaactgtc 901
DB   841 gaaatgtaacgacacagggcggtcaagccatcgccatcgtagtagtaataaaactgtc 900

QY   902 aatggaagcagcttttgataaagtaggtaccgtagctacgggtcaagtagtccatgcgcac 961
DB   901 aatggaagcagcttttgataaagtaggtaccgtagctacgggtcaagtagtccatgcgcac 960

QY   962 ggtggtgatgtaaaaggcttagtggtagcctcgatcgccatcggttagtgacttaacat 1021
DB   961 ggtggtgatgtaaaaggcttagtggtagcctcgatcgccatcggttagtgacttaacat 1020

QY   1022 ttgcttgatcagcattgtaattccttaaacatccgaaaagctactctgattaaacgatttatt 1081
DB   1021 ttgcttgatcagcattgtaattccttaaacatccgaaaagctactctgattaaacgatttatt 1080

QY   1082 aacggccatgcagttattaaaagaaatcacgaagctcaaaaggataatgattgtaaaataga 1141
DB   1081 aacggccatgcagttattaaaagaaatcacgaagctcaaaaggataatgattgtaaaataga 1140

QY   1142 cgcaaacacggcaagcgacacgacccagtagtactcagtgaggccatgtcatatgcacaggt 1201
DB   1141 cgcaaacacggcaagcgacacgacccagtagtactcagtgaggccatgtcatatgcacaggt 1200

QY   1202 cattttccaacgccttttggtagcagggcaacagcagcagcagcagcagcagcagcagcag 1261
DB   1201 cattttccaacgccttttggtagcagggcaacagcagcagcagcagcagcagcagcagcag 1260

QY   1262 ggtcttccgcccacagcagcagggcccaatctacaatcgatttggcttgatgcagcactct 1321
DB   1261 ggtcttccgcccacagcagcagggcccaatctacaatcgatttggcttgatgcagcactct 1320
```


Db 3481 gcaaatatcaagtcacacccctcaactcaaaaggtgaaacgcgtctttaattataaac 3540
Qy 3542 gacaaaatggtacggttaccttggcattaacaccacaagcggtcttaaaagcggcaaa 3601
Db 3543 gacaaaatggtacggttaccttggcattaacaccacaagcggtcttaaaagcggcaaa 3600
Qy 3602 agacccataacgacggtggtctctctattaaaaaccccccactggtagcgaacaataccaa 3661
Db 3603 agacccataacgacggtggtctctctattaaaaaccccccactggtagcgaacaataccaa 3660
Qy 3662 gtcggttgcgtgacggtggaagttggcaagtttaataatggtgtgtgtaggtgctggc 3721
Db 3663 gtcggttgcgtgacggtggaagttggcaagtttaataatggtgtgtgtaggtgctggc 3720
Qy 3722 attgatggcaactgcattaccagagatgaaattggctttactgggacttaattggctca 3781
Db 3723 attgatggcaactgcattaccagagatgaaattggctttactgggacttaattggctca 3780
Qy 3782 ctgtataaaagcaaacccaccctaaagcaagacggtattaacgcaggtggtataaaagatt 3841
Db 3783 ctgtataaaagcaaacccaccctaaagcaagacggtattaacgcaggtggtataaaagatt 3840
Qy 3842 accaacttcaatcaggtgagattggccaaacagccatgatgctgtgacagcgcgcaag 3901
Db 3843 accaacttcaatcaggtgagattggccaaacagccatgatgctgtgacagcgcgcaag 3900
Qy 3902 atttatgattaaaaacccgaacttgaaacaaaatcagcagtagctgcacaaacagcacaa 3961
Db 3903 atttatgattaaaaacccgaacttgaaacaaaatcagcagtagctgcacaaacagcacaa 3960
Qy 3962 aactcattacaggaattctcagtagcagatgaacaaaggttaataactttacggttagtaac 4021
Db 3963 aactcattacaggaattctcagtagcagatgaacaaaggttaataactttacggttagtaac 4020
Qy 4022 ccttactccagttatgacacctcaaaacacctctgatgtcatcacctttgcaggtgaaaaac 4081
Db 4023 ccttactccagttatgacacctcaaaacacctctgatgtcatcacctttgcaggtgaaaaac 4080
Qy 4082 ggcattaccacaaaggttaataaaaggtgtggtgcgttggcattgaccaaaccacaaaggc 4141
Db 4083 ggcattaccacaaaggttaataaaaggtgtggtgcgttggcattgaccaaaccacaaaggc 4140
Qy 4142 ttaaccacgctaagctgacggtgggttaaataatggcaaaagcatttgtcattgacagc 4201
Db 4143 ttaaccacgctaagctgacggtgggttaaataatggcaaaagcatttgtcattgacagc 4200
Qy 4202 caaaatggtcaaaataccatcacaggaactaagcaacactctagctaaattgtaccacatgat 4261
Db 4203 caaaatggtcaaaataccatcacaggaactaagcaacactctagctaaattgtaccacatgat 4260
Qy 4262 aaaggtgacgtacgcacacagaagcggcaataataataaaagacgaagcaaaaacccgt 4321
Db 4263 aaaggtgacgtacgcacacagaagcggcaataataataaaagacgaagcaaaaacccgt 4320
Qy 4322 gccccacgattgtgatgtgctaaagcgcaggtttaaacttgcaaggcaaatggtgaagcg 4381
Db 4323 gccccacgattgtgatgtgctaaagcgcaggtttaaacttgcaaggcaaatggtgaagcg 4380
Qy 4382 gttgactttgtctccacttatgacacgcgtcaactttgccgatggcaatgcccaccacgcct 4441
Db 4383 gttgactttgtctccacttatgacacgcgtcaactttgccgatggcaatgcccaccacgcct 4440
Qy 4442 aagtgacctatgatgacacaaagcaaacccagtaaaagttggtctatgatgtcaattggtgat 4501
Db 4443 aagtgacctatgatgacacaaagcaaacccagtaaaagttggtctatgatgtcaattggtgat 4500
Qy 4502 gatacaaccttgagttaaagaataaaaacttggcgtataaaacccacacattgaccaggt 4561
Db 4503 gatacaaccttgagttaaagaataaaaacttggcgtataaaacccacacattgaccaggt 4560
Qy 4562 actggcaggtgctataaattggcctaagcaatcaagctactggcgatgctgctgttc 4621
Db 4563 actggcaggtgctataaattggcctaagcaatcaagctactggcgatgctgctgttc 4620

Qy 4622 aagccagtgatctgttgcctcatctaaacaccccttatctggcgacatccaaactgccaaa 4681
Db 4623 aagccagtgatctgttgcctcatctaaacaccccttatctggcgacatccaaactgccaaa 4680
Qy 4682 gggcgaagcaagcgaacaactcagcaggtctatgtgtgctgtatggtgcaataaaggtcatc 4741
Db 4683 gggcgaagcaagcgaacaactcagcaggtctatgtgtgctgtatggtgcaataaaggtcatc 4740
Qy 4742 tatgacagtcaccgataacaagtactatcaagcccaaaaatgatggcacagttgataaaacc 4801
Db 4743 tatgacagtcaccgataacaagtactatcaagcccaaaaatgatggcacagttgataaaacc 4800
Qy 4802 aaagaattggccaaagacaaactggtgcccaagcccaaaacccagatggcacattggtct 4861
Db 4803 aaagaattggccaaagacaaactggtgcccaagcccaaaacccagatggcacattggtct 4860
Qy 4862 caaatgaatgcataatcagtcattaaacaaagaacaagttaaatgatgcataaaaaagcaa 4921
Db 4863 caaatgaatgcataatcagtcattaaacaaagaacaagttaaatgatgcataaaaaagcaa 4920
Qy 4922 ggcataatgaagacaaacgcctttgttaaaggacttgaaaaagcgccttctgtataacaaa 4981
Db 4923 ggcataatgaagacaaacgcctttgttaaaggacttgaaaaagcgccttctgtataacaaa 4980
Qy 4982 accaaaaacgcgcagtaactgttgaatttaaatgccttggcccaaacaccgcctgacc 5041
Db 4983 accaaaaacgcgcagtaactgttgaatttaaatgccttggcccaaacaccgcctgacc 5040
Qy 5042 ttctcaggggtacaggcacaaacgccttaaaaactggcgagactttgacacataaaaggt 5101
Db 5043 ttctcaggggtacaggcacaaacgccttaaaaactggcgagactttgacacataaaaggt 5100
Qy 5102 gggcaaacagacacccaataaagctaacgcgataataacatcgggtggtgagcaggtactgat 5161
Db 5103 gggcaaacagacacccaataaagctaacgcgataataacatcgggtggtgagcaggtactgat 5160
Qy 5162 ggccttcactgtccaaacttgccaaagacacccaatacttaacagcgtttaatgcaggtggc 5221
Db 5163 ggccttcactgtccaaacttgccaaagacacccaatacttaacagcgtttaatgcaggtggc 5220
Qy 5222 accaaaattgtagcaaaagcggtctctttttagactcaagcgggtcaagcgaacaaagcaac 5281
Db 5223 accaaaattgtagcaaaagcggtctctttttagactcaagcgggtcaagcgaacaaagcaac 5280
Qy 5282 acccctgtcgttaagtgccaatggcctggacacctgggtggcgaaggtcatcagtaattgggc 5341
Db 5283 acccctgtcgttaagtgccaatggcctggacacctgggtggcgaaggtcatcagtaattgggc 5340
Qy 5342 aaaggcacaagataccgcgcgtgccaatgtcaaacagtttaaacgaagtagcgaacttg 5401
Db 5343 aaaggcacaagataccgcgcgtgccaatgtcaaacagtttaaacgaagtagcgaacttg 5400
Qy 5402 ttgggtcttgtaattgctggttaatgataacgcgtgacggcaatcaggtaaaaactggccgac 5461
Db 5403 ttgggtcttgtaattgctggttaatgataacgcgtgacggcaatcaggtaaaaactggccgac 5460
Qy 5462 atcaaaaagaccacaaattcaggttcatctatcaacgcactgtcatcaaaagcagcgacg 5521
Db 5463 atcaaaaagaccacaaattcaggttcatctaaacgcactgtcatcaaaagcagcgacg 5520
Qy 5522 gtacttggcggttaaggttaataacgataccgaaaaacttgcactggtgggtatacaagt 5581
Db 5523 gtacttggcggttaaggttaataacgataccgaaaaacttgcactggtgggtatacaagt 5580
Qy 5582 ggcgtgataaagcggcaacgcgttaacgggatttaagaacatgtttgggtcaaaaccccaa 5641
Db 5583 ggcgtgataaagcggcaacgcgttaacgggatttaagaacatgtttgggtcaaaaccccaa 5640
Qy 5642 aaagatggcagcaaaaaaacccctgctccacttataacgcgcaggtgcacacacactat 5701
Db 5643 aaagatggcagcaaaaaaacccctgctccacttataacgcgcaggtgcacacacactat 5700

Matches 6258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 708 tgatcggtgaacgctcagtggtgcagtgcttatgtctcaaaaaaagataccaaaacatctcg 767
 Db 2 tgatcggtgaacgctcagtggtgcagtgcttatgtctcaaaaaaagataccaaaacatctcg 61
 Qy 768 caattgtgtaacaaacacagcaagagcgtcagagcgtcagagcgtcagagcgtgagcag 827
 Db 62 caattgtgtaacaaacacagcaagagcgtcagagcgtcagagcgtcagagcgtgagcag 121
 Qy 828 ccattgttattggtgtaaaatgctaacgcacagggcggtcgaagccatcgccatcggtgagta 887
 Db 122 ccattgttattggtgtaaaatgctaacgcacagggcggtcgaagccatcgccatcggtgagta 181
 Qy 888 gtaataaaactgtcaatggaagcagtttggataagataggtaccgatagtctacgggtcaag 947
 Db 182 gtaataaaactgtcaatggaagcagtttggataagataggtaccgatagtctacgggtcaag 241
 Qy 948 agtccatcgccatcggtggtgagatgaagcgtagtgatgctcgtcgtatgcccacgtgta 1007
 Db 242 agtccatcgccatcggtggtgagatgaagcgtagtgatgctcgtcgtatgcccacgtgta 301
 Qy 1008 gtgatgacttacatttctgtgatcagcatgtaacatccttaaacatccgaaggttactctga 1067
 Db 302 gtgatgacttacatttctgtgatcagcatgtaacatccttaaacatccgaaggttactctga 361
 Qy 1068 ttaacgatcttataacggtccatcgatgtattataaagaataacgaagctcaaaagataatg 1127
 Db 362 ttaacgatcttataacggtccatcgatgtattataaagaataacgaagctcaaaagataatg 421
 Qy 1128 atgtataaatatagacgcacaaacgcagcagcagccagctactgagtgaggagccatgt 1187
 Db 422 atgtataaatatagacgcacaaacgcagcagcagccagctactgagtgaggagccatgt 481
 Qy 1188 catatgcacagggctcattttccaaacgcttgggtacacggcgcaacagctaaaaagtgccct 1247
 Db 482 catatgcacagggctcattttccaaacgcttgggtacacggcgcaacagctaaaaagtgccct 541
 Qy 1248 attccttggaagtggtgttgcgcacacagccgagggccaaatctacaatcgctatttggtt 1307
 Db 542 attccttggaagtggtgttgcgcacacagccgagggccaaatctacaatcgctatttggtt 601
 Qy 1308 ctgatgaacatctagctcgttggagcgtatagcccttggcgaggttactcgtctcagc 1367
 Db 602 ctgatgaacatctagctcgttggagcgtatagcccttggcgaggttactcgtctcagc 661
 Qy 1368 tacaggcgagttatgcccctaggtcaaggttctgttgcactcagagtgataaatctcta 1427
 Db 662 tacaggcgagttatgcccctaggtcaaggttctgttgcactcagagtgataaatctcta 721
 Qy 1428 gaccggcctatacaccataacccagcacttagaccaccaagtttcaagccaccaataata 1487
 Db 722 gaccggcctatacaccataacccagcacttagaccaccaagtttcaagccaccaataata 781
 Qy 1488 cgaaggcggttccacttccacttggtagtaactctatacaacgttaaaatcatcaatgtcg 1547
 Db 782 cgaaggcggttccacttccacttggtagtaactctatacaacgttaaaatcatcaatgtcg 841
 Qy 1548 gtgcaggtgttataaaacacgctcgttcaatgtggcacagctagaagcggtgtggaagt 1607
 Db 842 gtgcaggtgttataaaacacgctcgttcaatgtggcacagctagaagcggtgtggaagt 901
 Qy 1608 gggctaaggagcgtagataattacttccaggtgtgatataacagctactgacgtataaaatag 1667
 Db 902 gggctaaggagcgtagataattacttccaggtgtgatataacagctactgacgtataaaatag 961
 Qy 1668 gtttggataataacttataacttaaaaggtgtgtgcagagaccaacgcattaacccgataata 1727
 Db 962 gtttggataataacttataacttaaaaggtgtgtgcagagaccaacgcattaacccgataata 1021
 Qy 1728 atatcgggtgtgttaaaagagcgtgataatagtggtctgaaaggttaaaacttctgtaaaactt 1787
 Db 1022 atatcgggtgtgttaaaagagcgtgataatagtggtctgaaaggttaaaacttctgtaaaactt 1081

Qy 1788 taaacaattcttactgaggtgaatacaactacattataatgccacaaccacagttaaggtag 1847
 Db 1082 taaacaattcttactgaggtgaatacaactacattataatgccacaaccacagttaaggtag 1141
 Qy 1848 gtagtagtagtagtactacagctgaattatigagtgaagtttaacctttaccagccca 1907
 Db 1142 gtagtagtagtagtactacagctgaattatigagtgaagtttaacctttaccagccca 1201
 Qy 1908 atacaggcagtcacaagcaagaacacgctctatggcggttaattgggtgaagtttacta 1967
 Db 1202 atacaggcagtcacaagcaagaacacgctctatggcggttaattgggtgaagtttacta 1261
 Qy 1968 ataatacagaacaaacacagcagcaatcgccactactcgtattaccagagataaaattggtc 2027
 Db 1262 ataatacagaacaaacacagcagcaatcgccactactcgtattaccagagataaaattggtc 1321
 Qy 2028 ttgctcagatggttgagtgttatgataaaacacagcaccatttttgataaaaaaacactta 2087
 Db 1322 ttgctcagatggttgagtgttatgataaaacacagcaccatttttgataaaaaaacactta 1381
 Qy 2088 aagtggtgtagtggttgcaattaccatagacaatggcattgatgcaggttaataaaagatca 2147
 Db 1382 aagtggtgtagtggttgcaattaccatagacaatggcattgatgcaggttaataaaagatca 1441
 Qy 2148 gtaattcttgcacaaaggtagcggtgctaacggtcggtttaccatcgcaacagctcaaaagcg 2207
 Db 1442 gtaattcttgcacaaaggtagcggtgctaacggtcggtttaccatcgcaacagctcaaaagcg 1501
 Qy 2208 ccaagctcacttttaaacgcagcgtgcacagctgctacacacttactgaaatcatcagttg 2267
 Db 1502 ccaagctcacttttaaacgcagcgtgcacagctgctacacacttactgaaatcatcagttg 1561
 Qy 2268 atgttaagtggtgcaattgttaccgcccacacttaacaacttggcggtgaaacacccagcg 2327
 Db 1562 atgttaagtggtgcaattgttaccgcccacacttaacaacttggcggtgaaacacccagcg 1621
 Qy 2328 ttaacagtgatggcactagtgataaaatttagttaaagggttagtggtacgacaataagct 2387
 Db 1622 ttaacagtgatggcactagtgataaaatttagttaaagggttagtggtacgacaataagct 1681
 Qy 2388 tagttaccgcgcgaacatttgcgaagctctataatgaagtcaatcgacggtgcagtg 2447
 Db 1682 tagttaccgcgcgaacatttgcgaagctctataatgaagtcaatcgacggtgcagtg 1741
 Qy 2448 ctctacaaggtttaccggttaaagaagacgagtgatgacgcccacgctatcacggtgg 2507
 Db 1742 ctctacaaggtttaccggttaaagaagacgagtgatgacgcccacgctatcacggtgg 1801
 Qy 2508 ctacaagatcacgacaaaaaatccggcgagtcagcatctttaaactcaaaaggttaaaaaag 2567
 Db 1802 ctacaagatcacgacaaaaaatccggcgagtcagcatctttaaactcaaaaggttaaaaaag 1861
 Qy 2568 gtctaaaggttgcctaccacaaaaagatgtagcgtttacgttggcggttagcacaagatagcg 2627
 Db 1862 gtctaaaggttgcctaccacaaaaagatgtagcgtttacgttggcggttagcacaagatagcg 1921
 Qy 2628 gtctgacatttggcaaaagcaccctaaacaacgtagggttgaactgttaagataccaacg 2687
 Db 1922 gtctgacatttggcaaaagcaccctaaacaacgtagggttgaactgttaagataccaacg 1981
 Qy 2688 aacaaatccaagtcggtgctaatggcattaaatttactaatgtgaatggtagtaataccag 2747
 Db 1982 aacaaatccaagtcggtgctaatggcattaaatttactaatgtgaatggtagtaataccag 2041
 Qy 2748 gtactggcatttgcacaaatccgctcattaccagagataaaattggcttctggtgtctg 2807
 Db 2042 gtactggcatttgcacaaatccgctcattaccagagataaaattggcttctggtgtctg 2101
 Qy 2808 atggtgcagttgtacacaaacacaccccttcttctgtaagacaaagctacaagttggcaatg 2867
 Db 2102 atggtgcagttgtacacaaacacaccccttcttctgtaagacaaagctacaagttggcaatg 2161

QY 2868 ttaagattaccacactggcattaaagcagggtggtaagccatcacagggtgtccccaa 2927
 Db 2162 ttaagattaccacactggcattaaagcagggtggtaagccatcacagggtgtccccaa 2221
 QY 2928 cactgctagcattgccgataaagtagccgcaacatagaactgggcaatacaatccaa 2987
 Db 2222 cactgctagcattgccgataaagtagccgcaacatagaactgggcaatacaatccaa 2281
 QY 2988 acaagacaaaatccacgctgcagcattaaatgatataataatcacaggtttaacctaa 3047
 Db 2282 acaagacaaaatccacgctgcagcattaaatgatataataatcacaggtttaacctaa 2341
 QY 3048 aaaaataacaaacccattgactttgtctccacttatgaactttgactttgccaatg 3107
 Db 2342 aaaaataacaaacccattgactttgtctccacttatgaactttgactttgccaatg 2401
 QY 3108 gcaatgccaccacccgacagtaaacccatgataccgtaacaaaacagtaagtggat 3167
 Db 2402 gcaatgccaccacccgacagtaaacccatgataccgtaacaaaacagtaagtggat 2461
 QY 3168 atgattgtaagtggatgatacaaacattcatataacaggcactgatgacaataaaaac 3227
 Db 2462 atgattgtaagtggatgatacaaacattcatataacaggcactgatgacaataaaaac 2521
 QY 3228 ttggcgtcaaaacccaaactgaacaaaacagtgacttaagtgaatacagcaactaa 3287
 Db 2522 ttggcgtcaaaacccaaactgaacaaaacagtgacttaagtgaatacagcaactaa 2581
 QY 3288 ttaattgtaacttagtgatgaagatgcccctgtttaacgccaagacatgcgcgaatac 3347
 Db 2582 ttaattgtaacttagtgatgaagatgcccctgtttaacgccaagacatgcgcgaatac 2641
 QY 3348 taaacccctagcaagaaatcaacaccaaagccagcagacacccgcccctacaa 3407
 Db 2642 taaacccctagcaagaaatcaacaccaaagccagcagacacccgcccctacaa 2701
 QY 3408 cctttacgttataaagtagatgaaaataatagtctgacgccaacgcacacacgc 3467
 Db 2702 cctttacgttataaagtagatgaaaataatagtctgacgccaacgcacacacgc 2761
 QY 3468 tgggtcaaaagcagcaaaataacagtcacacccctaaacacacacaggtgaaacggtc 3527
 Db 2762 tgggtcaaaagcagcaaaataacagtcacacccctaaacacacacaggtgaaacggtc 2821
 QY 3528 ttaataataaaacgcacaaaatgtagcgttaccttggcattacacacacagcggtc 3587
 Db 2822 ttaataataaaacgcacaaaatgtagcgttaccttggcattacacacacagcggtc 2881
 QY 3588 ttaaagccggcaaaagcaccctaaacagcgggtggtgtgtattataaaacccactggt 3647
 Db 2882 ttaaagccggcaaaagcaccctaaacagcgggtggtgtgtattataaaacccactggt 2941
 QY 3648 gcgaacaaatccagtcggtgctgagtcggtgaaagttggcaggttaataatggtg 3707
 Db 2942 gcgaacaaatccagtcggtgctgagtcggtgaaagttggcaggttaataatggtg 3001
 QY 3708 ttgtaggtgctggtgattgagtgcaacactgcattaccagagtgaaatggctttactg 3767
 Db 3002 ttgtaggtgctggtgattgagtgcaacactgcattaccagagtgaaatggctttactg 3061
 QY 3768 ggaactaatgctcacttgataaaagcaaacccacccctaaagcaaaagcggcattaacgag 3827
 Db 3062 ggaactaatgctcacttgataaaagcaaacccacccctaaagcaaaagcggcattaacgag 3121
 QY 3828 gtggtaaaagattaccacattcaatcaggtgagattgccaacacagccatgatgctg 3887
 Db 3122 gtggtaaaagattaccacattcaatcaggtgagattgccaacacagccatgatgctg 3181
 QY 3888 tgacagggcggaagttatgatttaaaccggaactgaaacaaatcagcagctactg 3947
 Db 3182 tgacagggcggaagttatgatttaaaccggaactgaaacaaatcagcagctactg 3241
 QY 3948 ccaaaacagcacaacacattcatcacgaattctcagtagcagatgaacaaaggtataact 4007

Db 3242 ccaaaacagcacaacacattcatcacgaattctcagtagcagatgaacaaaggtataact 3301
 QY 4008 ttacggttagtaaaccttactccagttatgacacactcaaaagacacctgacttcaact 4067
 Db 3302 ttacggttagtaaaccttactccagttatgacacactcaaaagacacctgacttcaact 3361
 QY 4068 ttgacggtgaaacagggcattaccaccaaggtataaataaagggtgtggtgctggtgacattg 4127
 Db 3362 ttgacggtgaaacagggcattaccaccaaggtataaataaagggtgtggtgctggtgacattg 3421
 QY 4128 accaaaccaaaggcttaaccacgctgaagctgacogtgggttaataataatggcaaaagca 4187
 Db 3422 accaaaccaaaggcttaaccacgctgaagctgacogtgggttaataataatggcaaaagca 3481
 QY 4188 ttgtcattgacagcaaaatggtcaaaataccatcacaggtactgaagcaacactctagcta 4247
 Db 3482 ttgtcattgacagcaaaatggtcaaaataccatcacaggtactgaagcaacactctagcta 3541
 QY 4248 atgtttaccaatgataaaggtagcgtacgcacacagaacagggcaataataatcaaaagacg 4307
 Db 3542 atgtttaccaatgataaaggtagcgtacgcacacagaacagggcaataataatcaaaagacg 3601
 QY 4308 aagacaaaacccgtgcccagcattgttgatgtgctgaagcagcagctttaaacttgcag 4367
 Db 3602 aagacaaaacccgtgcccagcattgttgatgtgctgaagcagcagctttaaacttgcag 3661
 QY 4368 gcaatgggtgaagcgttgactttgtctccacttatgacacogtcaactttgcggtggca 4427
 Db 3662 gcaatgggtgaagcgttgactttgtctccacttatgacacogtcaactttgcggtggca 3721
 QY 4428 atgcaacacccgtgaaggtgacactatgacacagcaaaacccagtaaaagtgtctatg 4487
 Db 3722 atgcaacacccgtgaaggtgacactatgacacagcaaaacccagtaaaagtgtctatg 3781
 QY 4488 atgcaatgtgatatacaacattgaagttaaagataaaaacttggcgttaaaaccca 4547
 Db 3782 atgcaatgtgatatacaacattgaagttaaagataaaaacttggcgttaaaaccca 3841
 QY 4548 ccacattgcccaggtactggcagcaggtgtcgttaataataatgtcccttaagcaatacagctactg 4607
 Db 3842 ccacattgcccaggtactggcagcaggtgtcgttaataataatgtcccttaagcaatacagctactg 3901
 QY 4608 gggatgcgcttgcagcagcaggtatcgttgcctatcaaacacttatcttgcgagca 4667
 Db 3902 gggatgcgcttgcagcagcaggtatcgttgcctatcaaacacttatcttgcgagca 3961
 QY 4668 tccaaactgcgaaggggcaagccaagcgaacaaactcagcaggtatgtggatgctgatg 4727
 Db 3962 tccaaactgcgaaggggcaagccaagcgaacaaactcagcaggtatgtggatgctgatg 4021
 QY 4728 gcaataaggtcatctatgacagtcacgataacagtaactactatcaagcccaaaaatgagga 4787
 Db 4022 gcaataaggtcatctatgacagtcacgataacagtaactactatcaagcccaaaaatgagga 4081
 QY 4788 cagttgataaaacccaaagaggttgcgaagcaaaactggtgcgcccaagccccccccag 4847
 Db 4082 cagttgataaaacccaaagaggttgcgaagcaaaactggtgcgcccaagccccccccag 4141
 QY 4848 atggcacattggtcctaataatgaaatgcaaatcagtcattaaacaaagcaacagtaaaatgag 4907
 Db 4142 atggcacattggtcctaataatgaaatgcaaatcagtcattaaacaaagcaacagtaaaatgag 4201
 QY 4908 ccaataaaagcagcagcattcaatgaagcaaacgcctttgttaaagacttgaaagacgg 4967
 Db 4202 ccaataaaagcagcagcattcaatgaagcaaacgcctttgttaaagacttgaaagacgg 4261
 QY 4968 cttctgatacaaaaacccaaacacgcgcagtaactgtgggtgatttaaaatgcggttgcgc 5027
 Db 4262 cttctgatacaaaaacccaaacacgcgcagtaactgtgggtgatttaaaatgcggttgcgc 4321
 QY 5028 aaacacgcctgacatttgcaggggatacagcacaacgcgttaaaaaacttggcgagactt 5087

Dbb 4322 aaacacccgctgacaccttgcaggggatacaggcgcaacacggctataaaaaactggcgagactt 4381
Qy 5088 tgacctcaaaagtgggcaaacagacacccaataagctaaacgataataaacatcgtgtgg 5147
Dbb 4382 tgacctcaaaagtgggcaaacagacacccaataagctaaacgataataaacatcgtgtgg 4441
Qy 5148 tagcaggtactgattggttcaactgcaaaacttgcaaaagacactcaacaaactttaacagcg 5207
Dbb 4442 tagcaggtactgattggttcaactgcaaaacttgcaaaagacactcaacaaactttaacagcg 4501
Qy 5208 ttaatgcaggtggcaccacaaattgatgacaaagcggtgtctttttagtagactcaagcggtc 5267
Dbb 4502 ttaatgcaggtggcaccacaaattgatgacaaagcggtgtctttttagtagactcaagcggtc 4561
Qy 5268 aagccaaagcaaacacccctgtgtaagtgcgaatggctggacactgggtggcgaaggtca 5327
Dbb 4562 aagccaaagcaaacacccctgtgtaagtgcgaatggctggacactgggtggcgaaggtca 4621
Qy 5328 tcagtaattggtggcaaaaggcaaaaagataccgacgctgccaatgatacaacagttaaaag 5387
Dbb 4622 tcagtaattggtggcaaaaggcaaaaagataccgacgctgccaatgatacaacagttaaaag 4681
Qy 5388 aagtacgcaactgttggtctgtgtaagtctgtaattgataacgctgacgccaatcagg 5447
Dbb 4682 aagtacgcaactgttggtctgtgtaagtctgtaattgataacgctgacgccaatcagg 4741
Qy 5448 taacattgcgcgacatcaaaaaagacccaatttcaggttcatcatcaaacgcgactgtca 5507
Dbb 4742 taacattgcgcgacatcaaaaaagacccaatttcaggttcatcatcaaacgcgactgtca 4801
Qy 5508 tcaagcaggcacggttacttggcgttaaaagtataacgataccgcaaaaaacttgccactg 5567
Dbb 4802 tcaagcaggcacggttacttggcgttaaaagtataacgataccgcaaaaaacttgccactg 4861
Qy 5568 gtgtatacagtggtggctgataaagacgcaacgctaacgcgatttaagcaatggtt 5627
Dbb 4862 gtgtatacagtggtggctgataaagacgcaacgctaacgcgatttaagcaatggtt 4921
Qy 5628 gggctcaaaaccccaaaagatggcagcaaaaaagccctgctgcacacttataaccccgag 5687
Dbb 4922 gggctcaaaaccccaaaagatggcagcaaaaaagccctgctgcacacttataaccccgag 4981
Qy 5688 gtcagaccaactatttgacaaacaccccgacagaagccattgacagaataatgaacaag 5747
Dbb 4982 gtcagaccaactatttgacaaacaccccgacagaagccattgacagaataatgaacaag 5041
Qy 5748 gtatccgctcttccatgtcaacgatggcaatcaagaccctgtgtacaaaggcgtaacg 5807
Dbb 5042 gtaaccgctcttccatgtcaacgatggcaatcaagaccctgtgtacaaaggcgtaacg 5101
Qy 5808 gcattgactcaagtgcctcaggcgaagcactcagtgggcgataggtttccaggcccaaggcag 5867
Dbb 5102 gcattgactcaagtgcctcaggcgaagcactcagtgggcgataggtttccaggcccaaggcag 5161
Qy 5868 atggtgaagcccgctgtgccatagggagaaaaacccaaagcggaaccaactccatcgcca 5927
Dbb 5162 atggtgaagcccgctgtgccatagggagaaaaacccaaagcggaaccaactccatcgcca 5221
Qy 5928 tcggtgataacgcacaagccagcgcatcaatccatcgccatcggtacaggcaatgtgg 5987
Dbb 5222 tcggtgataacgcacaagccagcgcatcaatccatcgccatcggtacaggcaatgtgg 5281
Qy 5988 tagcaggttaagcactctggtccatcgccgacccaagcactgttaaggctgataacagtt 6047
Dbb 5282 tagcaggttaagcactctggtccatcgccgacccaagcactgttaaggctgataacagtt 5341
Qy 6048 acagttgggttaataacaaacagtttaaccgatgccaactcaaacggatgcttttgggtgg 6107
Dbb 5342 acagttgggttaataacaaacagtttaaccgatgccaactcaaacggatgcttttgggtgg 5401
Qy 6108 gcaataacatcacctgacccaagaaagtaactcgttgccttaggttcaaacctgccatca 6167
Dbb 5402 gcaataacatcacctgacccaagaaagtaactcgttgccttaggttcaaacctgccatca 5461

Qy 6168 gtgcaggcacacagcagcgacacaaagccaaaaatctgcagcgcacagcaggtacaacca 6227
Dbb 5462 gtgcaggcacacagcagcgacacaaagccaaaaatctgcagcgcacagcaggtacaacca 5521
Qy 6228 ccaacagcaggtgcacaacggtaagcgtttaaagccttgcctggacaaacggcgttgggtgg 6287
Dbb 5522 ccaacagcaggtgcacaacggtaagcgtttaaagccttgcctggacaaacggcgttgggtgg 5581
Qy 6288 tctccgtgggtgcctcaggtgctgaacccgctatccaaaaatgtgcagcaggtgaggtca 6347
Dbb 5582 tctccgtgggtgcctcaggtgctgaacccgctatccaaaaatgtgcagcaggtgaggtca 5641
Qy 6348 gtgcaccagcacccgatcggttcaatgtgaccagttgtacaaaagcccccacaaacattg 6407
Dbb 5642 gtgcaccagcacccgatcggttcaatgtgaccagttgtacaaaagcccccacaaacattg 5701
Qy 6408 ccaacgcaacaaatgagcttgaccatcgtatccacacaaagcaaaaatgaaggccaatgag 6467
Dbb 5702 ccaacgcaacaaatgagcttgaccatcgtatccacacaaagcaaaaatgaaggccaatgag 5761
Qy 6468 ggatttcatcagcagtgatggcgctccatgcacaaagcctacattcctgcagatcca 6527
Dbb 5762 ggatttcatcagcagtgatggcgctccatgcacaaagcctacattcctgcagatcca 5821
Qy 6528 tggttaccgggggtatttgcaccccaacacggtcaaggtgcggtggcagtgaggactgtcga 6587
Dbb 5822 tggttaccgggggtatttgcaccccaacacggtcaaggtgcggtggcagtgaggactgtcga 5881
Qy 6588 agctgcggaataatggtcaatgggtatttataaatacaatgggttcagccgatacccaagcc 6647
Dbb 5882 agctgcggaataatggtcaatgggtatttataaatacaatgggttcagccgatacccaagcc 5941
Qy 6648 atgtagggcgcgagtggtgcaggtttcacttttaagccataaatcgcaagattttac 6707
Dbb 5942 atgtagggcgcgagtggtgcaggtttcacttttaagccataaatcgcaagattttac 6001
Qy 6708 ttaaaaaatcaatctcaccatagttgtataaaacagcagcatcagtcagtcataactgat 6767
Dbb 6002 ttaaaaaatcaatctcaccatagttgtataaaacagcagcatcagtcataactgat 6061
Qy 6768 gctgattgtttttatcaattaaacattttaccgctcaagtgattcttccaccatgac 6827
Dbb 6062 gctgattgtttttatcaattaaacattttaccgctcaagtgattcttccaccatgac 6121
Qy 6828 caaatcgccattgatacaggtataaacttatgtgataaattttatcaatgtagttgttaga 6887
Dbb 6122 caaatcgccattgatacaggtataaacttatgtgataaattttatcaatgtagttgttaga 6181
Qy 6888 tatggttaaaattgtgccattgacccaaaaaatgaccgattttatccgaaaaattttctgatt 6947
Dbb 6182 tatggttaaaattgtgccattgacccaaaaaatgaccgattttatccgaaaaattttctgatt 6241
Qy 6948 atgatacgttgacactgca 6965
Dbb 6242 atgatacgttgacactgca 6259

RESULT 5

AAF59103

ID AAF59103 standard; DNA; 6144 BP.

.XX

AC AAF59103;

XX AC

XX AC

DT 24-APR-2001 (first entry)

XX DT

XX DT

DE M. catarrhalis strain 4223 genomic 200kDa coding sequence SEQ ID NO:6.

XX DE

XX DE

KW Moraxella catarrhalis strain 4223; major outer membrane protein;

XX KW

KW Moraxella catarrhalis strain 4223; major outer membrane protein;

XX KW

KW Moraxella catarrhalis strain 4223; major outer membrane protein;

XX KW

OS Moraxella catarrhalis.

Db 5881 gagcttgaccatcgtatccaccacaaagaaataagcccaatgcagggtatttcacagcg 5940
Qy 6482 atggcgatggcgctccatgcccacagctacattctctgcagatccatggttacgggggt 6541
Db 5941 atggcgatggcgctccatgcccacagctacattctctgcagatccatggttacgggggt 6000
Qy 6542 atggccaccacacagctcaaggtgcggtgcagtgagggtgcagctgcggatgcggataat 6601
Db 6001 atggccaccacacagctcaaggtgcggtgcagtgagggtgcagctgcggatgcggataat 6060
Qy 6602 ggtcaatgggtatttaaaatcaatggttcagccgatacccaaggccatgtaggggcgga 6661
Db 6061 ggtcaatgggtatttaaaatcaatggttcagccgatacccaaggccatgtaggggcgga 6120
Qy 6662 gttggtgcaggttttccacttttaa 6685
Db 6121 gttggtgcaggttttccacttttaa 6144

RESULT 6
AAF59101
ID AAF59101 standard; DNA: 5979 BP.
XX
AC AAF59101;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa coding sequence #2.
XX
KW Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection; ds.
XX
OS Moraxella catarrhalis.
XX
PN W0200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-CA00870.
XX
PR 27-JUL-1999; 990S-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
PT WPI; 2001-159722/16.
DR P-PSDB; AAB69133.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
PS Example 3; Fig 2A-W; 247pp; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa coding sequence,
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 5979 BP; 1974 A; 1360 C; 1355 G; 1290 T; 0 other;

Query Match 85.8%; Score 5979; DB 22; Length 5979;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 707 gtgacgggtgcacagctcagtggtgagtgcttatgtctcaaaaaaagataccaaacatc 766
Db 1 gtgacgggtgcacagctcagtggtgagtgcttatgtctcaaaaaaagataccaaacatc 60
Qy 767 gcaattggtgaacaaaaccagcgaagcgtcaggcactccaaggcgagcggtgacga 826
Db 61 gcaattggtgaacaaaaccagcgaagcgtcaggcactccaaggcgagcggtgacga 120
Qy 827 gccattgctattggtgaaaaatgctaaagcagcggtgcaagccatcgccatcggtagt 886
Db 121 gccattgctattggtgaaaaatgctaaagcagcggtgcaagccatcgccatcggtagt 180
Qy 887 agtaataaaactgtcaatggaagcagtttgataagataggtaccgagtgtacgggtcaa 946
Db 181 agtaataaaactgtcaatggaagcagtttgataagataggtaccgagtgtacgggtcaa 240
Qy 947 gagtccatgccatcggtggtgataagagctagtgatgctcgtcgtcgtcgtcgtcgt 1006
Db 241 gagtccatgccatcggtggtgataagagctagtgatgctcgtcgtcgtcgtcgtcgt 300
Qy 1007 agtgatgacttacatttcttgatcagcatggttaactcctaaccatccgaaggttactctg 1066
Db 301 agtgatgacttacatttcttgatcagcatggttaactcctaaccatccgaaggttactctg 360
Qy 1067 attaacgatcttataacggtccatgcagtgatataaaagaaatacagaagctcaaaagataat 1126
Db 361 attaacgatcttataacggtccatgcagtgatataaaagaaatacagaagctcaaaagataat 420
Qy 1127 gatgaaaaatagacgacacaaaccgacgacacccagcagctactcagtgaggagccatg 1186
Db 421 gatgaaaaatagacgacacaaaccgacgacacccagcagctactcagtgaggagccatg 480
Qy 1187 tcatatgcacaggttcattttccaaagcctttgttacacgggcaacagctcaaaagtgc 1246
Db 481 tcatatgcacaggttcattttccaaagcctttgttacacgggcaacagctcaaaagtgc 540
Qy 1247 tattccttgagtggtggtcttgccgacacagccagggccaaatctacaatcgctattggt 1306
Db 541 tattccttgagtggtggtcttgccgacacagccagggccaaatctacaatcgctattggt 600
Qy 1307 tctgacgacacatcagctcgttggtgagcagatagcccttggtcaggtactcgtcgtcag 1366
Db 601 tctgacgacacatcagctcgttggtgagcagatagcccttggtcaggtactcgtcgtcag 660
Qy 1367 ctacaggcaggtattgcccctaggtcaaggttctgttgcactcagagtgataataattct 1426
Db 661 ctacaggcaggtattgcccctaggtcaaggttctgttgcactcagagtgataataattct 720
Qy 1427 agaccggcctatacacaaataaccacggcactagaccccaagtttcaagccaccaataat 1486
Db 721 agaccggcctatacacaaataaccacggcactagaccccaagtttcaagccaccaataat 780
Qy 1487 acgaagcggttcacatttccattggttagtaactctatcaacgtataaatcatcaatc 1546
Db 781 acgaagcggttcacatttccattggttagtaactctatcaacgtataaatcatcaatc 840
Qy 1547 ggtcaggtggttaataaaaccgagtcggtcaatgtggcacagctagaagcggtggtgaag 1606
Db 841 ggtcaggtggttaataaaaccgagtcggtcaatgtggcacagctagaagcggtggtgaag 900
Qy 1607 tgggctaaggagcgtagaattacttttcaggggtgatgatacagctactgcgtataaata 1666
Db 901 tgggctaaggagcgtagaattacttttcaggggtgatgatacagctactgcgtataaata 960
Qy 1667 ggttgataatactttaactattaaaggtggtgcagagacccaacgataacgataat 1726
Db 961 ggttgataatactttaactattaaaggtggtgcagagacccaacgataacgataat 1020

Db 3181 gtgacaggcggaagatttatgatttaaaacccgaacttgaaaaacaaataacagcagtact 3240
Qy 3947 gccaaaacagcacaaaactcattacagaaattctcagtagcagatgacaggtaataaac 4006
Db 3241 gccaaaacagcacaaaactcattacagaaattctcagtagcagatgacaggtaataaac 3300
Qy 4007 ttacggttagtaacccttactccagttatgacacctcacaagacctctgtatgtcateacc 4066
Db 3301 ttacggttagtaacccttactccagttatgacacctcacaagacctctgtatgtcateacc 3360
Qy 4067 ttgcaagtgaaaaacgcatcattaccacaaagtgaaataaaagttggtgctgctggtggcatt 4126
Db 3361 ttgcaagtgaaaaacgcatcattaccacaaagtgaaataaaagttggtgctgctggtggcatt 3420
Qy 4127 gacaaaccaaagccttaaccacgcttaagctgacccgtgggtggttaataataatggcacaagggc 4186
Db 3421 gacaaaccaaagccttaaccacgcttaagctgacccgtgggtggttaataataatggcacaagggc 3480
Qy 4187 attgtcattgacagccaaaattggtcaaaaataccatcacagagactaaacacactctagct 4246
Db 3481 attgtcattgacagccaaaattggtcaaaaataccatcacagagactaaacacactctagct 3540
Qy 4247 aatgttaccaatgataaaggtagcgtacgcacacacagacagggcgaataataatacaagac 4306
Db 3541 aatgttaccaatgataaaggtagcgtacgcacacacagacagggcgaataataatacaagac 3600
Qy 4307 gaagacaaaacccgtgcgcagcattgttgatgtgtaagcgcagggctttaacttgcga 4366
Db 3601 gaagacaaaacccgtgcgcagcattgttgatgtgtaagcgcagggctttaacttgcga 3660
Qy 4367 ggcgaatggtgaagcgttgactttgtctcaacttatagacacgtcaactttgcccagtgcc 4426
Db 3661 ggcgaatggtgaagcgttgactttgtctcaacttatagacacgtcaactttgcccagtgcc 3720
Qy 4427 aatgccaccacgttaaggtgacctatgatgacacagcaaaaacccagtaagtggtctat 4486
Db 3721 aatgccaccacgttaaggtgacctatgatgacacagcaaaaacccagtaagtggtctat 3780
Qy 4487 gatgtcaatgtggatgatacaaccattgaagttaaagataaaaacacttggcgtaaaaacc 4546
Db 3781 gatgtcaatgtggatgatacaaccattgaagttaaagataaaaacacttggcgtaaaaacc 3840
Qy 4547 accacattgaccagttactggcacagggtgctaaataaaattgcccctaagcaatacaagctact 4606
Db 3841 accacattgaccagttactggcacagggtgctaaataaaattgcccctaagcaatacaagctact 3900
Qy 4607 ggcgatgcgcttgcaaggccagtgatcgttgctcatctaaacacacttatctgcccag 4666
Db 3901 ggcgatgcgcttgcaaggccagtgatcgttgctcatctaaacacacttatctgcccag 3960
Qy 4667 atccaaactgccaaaggggcgaagcgaagcgaacaaactcagcaggtatgttgatgtctgat 4726
Db 3961 atccaaactgccaaaggggcgaagcgaagcgaacaaactcagcaggtatgttgatgtctgat 4020
Qy 4727 ggcgaataaggtcatctatgacagttaccgataaacaagtactatcaagccaaaataatgatggc 4786
Db 4021 ggcgaataaggtcatctatgacagttaccgataaacaagtactatcaagccaaaataatgatggc 4080
Qy 4787 acagttgataaaaccaaagaagttgccaaagacaaaactggtcgccaaagcccaacccca 4846
Db 4081 acagttgataaaaccaaagaagttgccaaagacaaaactggtcgccaaagcccaacccca 4140
Qy 4847 gatggcacattgggtcctcaaatgaaatgtaaatcagtcattcaacaagaacaaagtaaatgat 4906
Db 4141 gatggcacattgggtcctcaaatgaaatgtaaatcagtcattcaacaagaacaaagtaaatgat 4200
Qy 4907 gccataaaaagcaaggaatcctaataagacaaacgcctttgttaaggacttgaaaaagcc 4966
Db 4201 gccataaaaagcaaggaatcctaataagacaaacgcctttgttaaggacttgaaaaagcc 4260
Qy 4967 gcttctgatacaaaaaccaaagccgcagtaactggtggttgaatggttgaatggttgc 5026
Db 4261 gcttctgatacaaaaaccaaagccgcagtaactggtggttgaatggttgaatggttgc 4320

Qy 5027 caaacaccgtgacctttgcagggtgatacagggcacacagcgttaaaaaactggcgagact 5086
Db 4321 caaacaccgtgacctttgcagggtgatacagggcacacagcgttaaaaaactggcgagact 4380
Qy 5087 ttgaccattcaaaagtgggcaaacagacacccaataaagctaaacgataataaacaatcgggtg 5146
Db 4381 ttgaccattcaaaagtgggcaaacagacacccaataaagctaaacgataataaacaatcgggtg 4440
Qy 5147 gtacaggtactgagtctcacttgcacttgcacacttgcacaaagacacccaatactttaaagc 5206
Db 4441 gtacaggtactgagtctcacttgcacttgcacaaagacacccaatactttaaagc 4500
Qy 5207 gttaatcaggtggcaccacaaatgtgacaaaagcgtgtctttttagactcaagcgt 5266
Db 4501 gttaatcaggtggcaccacaaatgtgacaaaagcgtgtctttttagactcaagcgt 4560
Qy 5267 caagccaaagcaaacacccctgtctaaagtgcacaaatgggctggacctgggtggcaagctc 5326
Db 4561 caagccaaagcaaacacccctgtctaaagtgcacaaatgggctggacctgggtggcaagctc 4620
Qy 5327 atcagtaagtggcacaagcacaacaaagataccgacgtgcacaaatgtacaacagttaaac 5386
Db 4621 atcagtaagtggcacaagcacaacaaagataccgacgtgcacaaatgtacaacagttaaac 4680
Qy 5387 gaagtacgcaacttgttgggtcttggtaagtctgttaataatgataacgctgacggcaaatcag 5446
Db 4681 gaagtacgcaacttgttgggtcttggtaagtctgttaataatgataacgctgacggcaaatcag 4740
Qy 5447 gtaaacattgcccagacatacaaaaagaccccaaatcaggctcatcatctaaacgcaactgc 5506
Db 4741 gtaaacattgcccagacatacaaaaagaccccaaatcaggctcatcatctaaacgcaactgc 4800
Qy 5507 atcaagcagcagcgttacttggcgttaaggttaataacgataccgaaaaacttgcacact 5566
Db 4801 atcaagcagcagcgttacttggcgttaaggttaataacgataccgaaaaacttgcacact 4860
Qy 5567 ggtggtatcacagtggcggtggtataaagcagcgcacgcgttaacggttttaagcaatgtt 5626
Db 4861 ggtggtatcacagtggcggtggtataaagcagcgcacgcgttaacggttttaagcaatgtt 4920
Qy 5627 tgggtcaaaaaccccaaaaagatggcagcaaaaagccctgtctgcacattataacgcccga 5686
Db 4921 tgggtcaaaaaccccaaaaagatggcagcaaaaagccctgtctgcacattataacgcccga 4980
Qy 5687 ggtcagacccaactatttgaccaacaccccgcagaagccattgacagataataatgaacaa 5746
Db 4981 ggtcagacccaactatttgaccaacaccccgcagaagccattgacagataataatgaacaa 5040
Qy 5747 ggtatccgcttcttccatgtcaacgattggcaatcaagagcctgtgttacaagggcgtaac 5806
Db 5041 ggtatccgcttcttccatgtcaacgattggcaatcaagagcctgtgttacaagggcgtaac 5100
Qy 5807 ggcattgactcaagtgcctcagggcaagcactcagtggttaggtttccaggccaaaggca 5866
Db 5101 ggcattgactcaagtgcctcagggcaagcactcagtggttaggtttccaggccaaaggca 5160
Qy 5867 gatgtaagcccgcttggccataggcagacaaacccaagcagcaacccaatccatccgc 5926
Db 5161 gatgtaagcccgcttggccataggcagacaaacccaagcagcaacccaatccatccgc 5220
Qy 5927 atcgggtgataacgcacaaagccacggcgatcaatccatcgccatcggttacaggcaatgtg 5986
Db 5221 atcgggtgataacgcacaaagccacggcgatcaatccatcgccatcggttacaggcaatgtg 5280
Qy 5987 gtacaggtgaagcactctgttgcacatggcgaccccaagcactgtttaaggctgataacagct 6046
Db 5281 gtacaggtgaagcactctgttgcacatggcgaccccaagcactgtttaaggctgataacagct 5340
Qy 6047 tacagtggtggttaataacaacacagtttacagatgccaactcaaacagatgttttgggtg 6106
Db 5341 tacagtggtggttaataacaacacagtttacagatgccaactcaaacagatgttttgggtg 5400

Thu Sep 13 14:18:26 2001

Db 44767 gacacaaacacatcatacaaaaggggtgctcccaaacactgactgacattaccaacacagat 44826
QY 2951 agtagcgcaacatagaactgggcaataacataccaagacaaagacaaatccaacgctgacc 3010
Db 44827 aatacacacacacagaaacagacatgacgcccagggcaag---aatccaaagcgtgcc 44883
QY 3011 agcatataatgatatataatcagcgttttaacctataaaataataaacaacccattgac 3070
Db 44884 agtacaagatgctcaaacagcgttttaacctgcaaggaalggcaagcccggtgac 44943
QY 3071 ttgtctcaactatgacattgtgactttgccaatggcaatgccaccacgcgcacagta 3130
Db 44944 ttgtctacttatgataccgttcaacttggcaatggcgatggcaccactgccaagtt 45003
QY 3131 acc-----catgatccgctacaacaaacagtaaaagtgtatagtgaatgtgat 3184
Db 45004 acctataattataatgaaggcaaacacacagcacagtggtgtctataatgtcaatgtgat 45063
QY 3185 gatacaaccattcatctaacagcgactgatgacaataaaaaacttggcgctcaaaaccacc 3244
Db 45064 gatacaaccattcatctggcaataataatggttaataaaaaacttagcttaaaaccacc 45123
QY 3245 aaactgaacaaacaaagtgtaatgttaatacagaactaaactttaatgttaactctagt 3304
Db 45124 acattgacaaacaaagtgtcaaatggttaatgcaactaaatcttagtgccaac----- 45177
QY 3305 gatgaagatgccctgttaacgccaagacagatcgcgaaataactcaaacaccttagccaag 3364
Db 45178 aatggcgatgccctgttaacgccaaggttatggcgacaacttaaaactttagcgcgag 45237
QY 3365 gaaattcaacacacaaaggcagcgagacacgcgcctacaaacctttaccgttaaaag 3424
Db 45238 gaaattcaacacacaaaggcagcgagacacgcgcctacaaacctttaaagtcacaaaa 45297
QY 3425 gtatgaaataataatgctgtafgacgcaacgccctaccctgggtggtcaaaagacgca 3484
Db 45298 gacggtgcaaccaatgatgag-----acaccatcacccgggt----- 45336
QY 3485 aataatcaagtoaacacccctaacactaaaggtgaaacggcttaataataaacccgac 3544
Db 45337 aaaaatgcgtgacacccctaacattcaaaaggtgaaacggcttaataataaacccgac 45396
QY 3545 aaaaatgctacgtttacctttgcaattacacacacacacacacacacacacacacacac 3601
Db 45397 aagaatggtcacggttacctttgcaattacacacacacacacacacacacacacacacac 45456
QY 3602 agcacccctaaacgacggtggtgtgtctattataaaacccacacacacacacacacacac 3661
Db 45457 accactctaaacaaagatggctgtgtctattataaaacccacacacacacacacacac 45516
QY 3662 gtcggtcgtgagtggtgaggttggccaaaggttaataatggttgggtgaggtgctggc 3721
Db 45517 gtcggtgctgagtggtgaggttggccaaaggttaataatggttgggtgaggtgctggc 45576
QY 3722 atcgatggcaaacactgcattaccagagatgaaatggcctttactgggacttaatggctca 3781
Db 45577 atcgatggcaaacactgcattaccagagatgaaatggcctttactgggacttaatggctca 45636
QY 3782 ctgataaaagcaaacacccac 3841
Db 45637 ctgataaac 45696
QY 3842 accaacattcaatcaggtgagatggccaaacacacacacacacacacacacacacacacac 3901
Db 45697 accaacattcaatcaggtgagatggccaaacacacacacacacacacacacacacacac 45756
QY 3902 atttatgattataaac 3961
Db 45757 atttatgattataaac 45816
QY 3962 aactcattacacgaattctcagtagcagatgaacaaaggttaataactttcaggttagtaac 4021

Db 45817 aactcattacacgaattctcagtagcagatgaacaaaggttaataactttacggttagtaac 45876
QY 4022 ccttactccagttatgacacactccaagacctctgatgcatcacctttgaggtgaaaaac 4081
Db 45877 ccttactccagttatgacacactccaagacctctgatgcatcacctttgaggtgaaaaac 45936
QY 4082 ggcattaccaccaagtaataaaaggtggtgctggtgggcatgacacaaacccaaagc 4141
Db 45937 ggcattaccaccaagtaataaaaggtggtgctggtgggcatgacacaaacccaaagc 45996
QY 4142 ttaaccacgctaaagctgacgctgggtggttaataataatggcaaaaggttcttgcacagc 4201
Db 45997 ttaaccacgctaaagctgacgctgggtggttaataataatggcaaaaggttcttgcacagc 46056
QY 4202 caaaatggtcaaaataccatcacagactaaagcaaacactctagcttaagtttaccaatgat 4261
Db 46057 aaagatggtcaaaataccatcacagactaaagcaaacactctagcttaagtttaccaatga- 46115
QY 4262 aaagtgtagctgacgacacacagacagggggcaatataatacaagacgacgaacaaacccgt 4321
Db 46116 ----tggtaggacgacgactaaagccaagggcttgcgataaacaccgacaaacccgt 46170
QY 4322 gccccaagcattgttgatgctaaagcgacggtctttaaacttgaaggaatgggtgaagcg 4381
Db 46171 gctgccagcattggctgatttaataatgcaggtttaaacttgcaggaatgggtacagcg 46230
QY 4382 gttgactttgtctccacttatgacacgcgtcaaacctttgcgcatggcaatggcacacgcgt 4441
Db 46231 gttgactttgtctccacttatgacacgtcaaacctttgcgcatggcaatggcacacgcgt 46290
QY 4442 aaggtgaactatgtagacacagcaaacacagtaaggtggtctatgatgcaatgtggat 4501
Db 46291 aaggtgaactatgtagacacagcaaacacagtaaggtggtctatgatgcaatgtggat 46350
QY 4502 gatacaacattgaaagttaaaagataaaaaacttgggttaaaaaacacacacacacacacac 4561
Db 46351 gatacaacattgaaagttaaaagataaaaaacttgggttaaaaaacacacacacacacac 46410
QY 4562 actggcacaggtgctcaataaaatttgcctaaagcaatcaagctactggcgatgctgtgct 4621
Db 46411 actggcacaggtgctcaataaaatttgcctaaagcaatcaagctactggcgatgctgtgct 46470
QY 4622 aagggcagtgatctgtgctcatctataaacacacacacacacacacacacacacacac 4681
Db 46471 aagggcagtgatctgtctcatctataaacacacacacacacacacacacacacacac 46530
QY 4682 ggggcaagccaagcgaaac 4741
Db 46531 ggggcaagccaagcgaaac 46590
QY 4742 tatgacagtacgatacaacagtaactatcaagccaaaaatgattggcacagttgataaaacc 4801
Db 46591 tatgacagtacgatacaacagtaactatcaagccaaaaatgattggcacagttgataaaacc 46650
QY 4802 aaagaagttgccaagac 4861
Db 46651 aaagaagttgccaagac 46710
QY 4862 caaatgaatgcaaatcagtcattacaagaagaagaataatgattgcaatgcaatgcaatgcaat 4921
Db 46711 caaatgaatgcaaatcagtcattacaagaagaagaataatgattgcaatgcaatgcaatgcaat 46770
QY 4922 ggcataatgaagac 4981
Db 46771 ggcataatgaagac 46830
QY 4982 accaaaaacgcagtaacttgggtgatttaaatgctgttcccaaacacacacacacacacac 5041
Db 46831 accaaaaacgcagtaacttgggtgatttaaatgctgttcccaaacacacacacacacacac 46890
QY 5042 ttgacggggatagacagac 5101
Db 46891 ttgacggggatagacagac 46950

Qy	2069	ttggataaaaaaacacttaaaaggggttagttgccaattaccatagacaatggcattgat	2120
Db	1513	cttgacaacgaaagctaaagtgtggcaacgacacctaaacagtggttagctgtt	1572
Qy	2129	gcaggtaataaaaaagatcagtaactctgccaaaagtagcagtgctaaacgacggttacc	2188
Db	1573	ataaacacactggtaataaacaatccaagtcggctgaatgcatataatttgcaca	1632
Qy	2189	atcgaaacagctcaaaagccgccaagcctacttttaaagcaggcgtggtcactagtcaca	2248
Db	1633	gtcgtataatagtgtgcaaatacctcagcaacagtcggcactgctgtattaccgaagag	1692
Qy	2249	ccctactgaaatatcagttgatgctaagagtggcaatgttaccgccccaacttacaaca--	2306
Db	1593	aaat1gggtt1ggtt1ggttactaatgat1gggtt1gataaacagcaccatt1ggataaa	1752
Qy	2307	-----ttggcgtgaaacaccagcagcttaacagtgatggcactagtgataaatttagt	2359
Db	1753	gaacgact1aaagtggctgcttgaaattaccacagatggtgattaaatgctgtaatt	1812
Qy	2360	gttaagggtagtggtacgaacaatagcttagttaccgcccgaacatttggcgaactatcta	2419
Db	1813	cacaaattaccggacttactaatgtagcaaatccgatgcggttaccatcaaacag	1872
Qy	2420	aalgaagtcaatcgaaacgctgacagtgctctacaaaagctttaccgttaagaagaagac	2479
Db	1873	ctcaagagcgcgaagcctactttaacgcgagggatggcatcagtatcattatagtaatac	1932
Qy	2480	gatgatgacg-----ccaaacgctatcacgctggct	2509
Db	1933	ggggatctagtgtatagtagtggaatatattaccaccccaactataacattagcgtgaaa	1992
Qy	2510	aaagatcagacaaaaaattccggcgcagtcagcatcttaaactcaaaaggttaaaaacggt	2569
Db	1993	accactaaagcttaacagtaattggccacagtggtataataaatttagt1ggttagtaatgc	2052
Qy	2570	ctaacggtgtctaccaaaaaagatgtaacggttaccttgggttagccaagatagcgggt	2629
Db	2053	catgatacaaatagcttagttaccgccaagatttggcagactatcaataaagtcaat	2112
Qy	2630	ctgacctatggcaaaagacaccttaaacacgatggcttgactgttaaaagataccaacgaa	2689
Db	2113	gaacggctgacagtgctctacaagctttaaag1cccaaaacggt1gataatagcaaac	2172
Qy	2690	caaatccaagtcggtgctaatggcattaaatttactaat1gtaatgggttagtaatccaggt	2749
Db	2173	gccatcacgctgggtaaagatacaaacggcaagaccttcaaaccttcaaacctcaaaaggt	2232
Qy	2750	actggcattgcaaataccgcctcgatata---ccagagataaaattggctttgtcgttct	2806
Db	2233	gaaacgggtgtaattatcagcccaatagagcagagtcagcttagcttttggcattgac	2292
Qy	2807	gatgg1gcag1tgatac-----aaacaaacttactctgatcaagacaagcta--caa	2857
Db	2293	caaaagtaatggctccacacgcctagctgacgctgggtagcgtataa1gtaatcga	2352
Qy	2858	gttggcaatgttaagattaccacactggcattaaacgcaggtggtaagccatcacaggg	2917
Db	2353	ttggttat1gagcaagtccctagcgtgacggttaacagcaccaaaaacatcattaaagga	2412
Qy	2918	ctgtcccaacactgctagcatttccgcatcaaa1tagccgcaacatagaaactgggcaat	2977
Db	2413	ttgtccccaacactgctagcatttccagtcgaagtccaggtggcgcgaacatagcaactgggcaat	2472
Qy	2978	acaatccaagacaaagacaaa1cccaacgctgcagcat1aatgatatt1aaataccaagc	3037
Db	2473	acaatcgaaagaaaagacaaa1ccaacgctgccagat1tgatgat1gcttaaat1gcagc	2532
Qy	3038	tttaacctaaaaataataacaacccccattgactttgtctccacttatgaactttgtac	3097
Db	2533	tttaacctaaaaataat1ggcaagacaaaagact1tgtctccact1at1gacact1gttgcac	2592

QY	3098	tttgccaatg	gcaatg	ccaccacg	ccacag	ttaacc	catgat	caggct	taacaaa	ccaggt	3155
DB	2593	ttttcgtatg	gcaatgcc	caccacg	ccacag	ttaact	tatgat	gaag	ccaatca	ccaggt	2652
QY	3158	aaagtgg	tatatgat	gtga	tggatg	ataca	aaccat	catcat	taacag	gcactgat	3217
DB	2653	aaagtgg	cgtatg	atgtg	gaatg	gagaa	aaccat	tgaa	ctgca	ggcgat	2712
QY	3218	aataaaaa	aaacttgg	cgtcaaa	accaccaa	acttgaa	caaaa	caagtgt	ctaattg	taataca	3277
DB	2713	aagaa	caacttgg	cgtcaaa	accatca	aactgac	cgaa	caagtact	aatgt	taatgca	2772
QY	3278	gaactaa	acttta	atgtta	actctag	tatgat	agaatg	cccttgt	taacg	ccaaagacatc	3337
DB	2773	ac-----	-tac	atttag	tacgcag	catgac	cttgc	cttgt	taagc	agcagt	2823
QY	3338	gccgaaa	atctaaa	accacctag	ccaagga	aaatcc	accaccaa	aaagcag	acagac	acac	3397
DB	2824	gcggcaa	lctaa	accacctag	ccggagg	aaatcc	accaccaa	aaagcag	acacac	acac	2883
QY	3398	gccttaca	aaactttac	cgttaaaa	aggtag	atgaaaa	taata	taactgt	gatgc	gcgaac	3457
DB	2884	gccttaca	aaactttac	cgttaaaa	aggtag	atgaaaa	taata	taactgt	gatgc	acacac	2943
QY	3458	gccatcc	cgcttga	gcgtca	aaagacg	caata	ataatca	agttca	accctta	ccactca	3517
DB	2944	gccatcc	cgcttga	gcgtca	aaagacg	caata	ataatca	agttca	accctta	ccactca	3003
QY	3518	gaaac	cggtctta	ataattaaa	accgca	aaaaatg	gtgtac	ggtttacc	tttgg	cattaa	3577
DB	3004	aaaa	cggtctt	gatattaaa	accgca	aaaaatg	gtgtac	ggtttacc	tttgg	cattaa	3063
QY	3578	acaag	cggtctta	aaagc	ggcaaa	agcac	---ctaa	acgag	gtgtgt	gtcttata	3634
DB	3064	caaag	cggtctt	aaagc	ggcgac	agcac	actctta	aaacaa	aatgg	cttgtctata	3123
QY	3635	aaccoc	acttgg	tago	gaacaa	atcc	aaagtgtgt	gtgtgt	gaagttg	gaagttg	3694
DB	3124	aaac	cgcttag	taacga	acaaatcc	caagtc	ggtgtgt	gtgtgt	gaagttg	gaagttg	3182
QY	3695	aataata	atgggtt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	3754
DB	3183	--taa	taattgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	3240
QY	3755	attgg	cttact	gggact	aatgg	ctcact	gtataa	agcaaa	ccccac	cttaagca	3814
DB	3241	attgg	cttact	gggact	aatgg	ctcact	gtataa	agcaaa	ccccac	cttaagca	3300
QY	3815	ggc	attaacg	aggtt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	3874
DB	3301	ggc	attaacg	aggtt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	gtgtgt	3360
QY	3875	agcc	atgat	ctgtgt	ga	caggcg	gaagatt	gtatt	taaaaa	accgaact	3934
DB	3361	agcc	atgat	ctgtgt	ga	caggcg	gaagatt	gtatt	taaaaa	accgaact	3420
QY	3935	atc	agcag	tactg	ccaaaa	acagca	caaaa	actctat	acagaa	ttctcag	3994
DB	3421	atc	agcag	tactg	ccaaaa	acagca	caaaa	actctat	acagaa	ttctcag	3480
QY	3995	caag	gtaata	actttac	ggttag	taac	cccttact	ccagtt	atgac	actcaag	4054
DB	3481	caag	gtaata	actttac	ggttag	taac	cccttact	ccagtt	atgac	actcaag	3540
QY	4055	gat	gtaata	caactttg	caggtg	aaaa	cggaatt	accac	caagtt	ataata	4114
DB	3541	gat	gtaata	caactttg	caggtg	aaaa	cggaatt	accac	caagtt	ataata	3600
QY	4115	cat	gtgg	catg	acaa	ccaaag	cttaac	ccac	gcctaa	gcgtgac	4174
DB	3601	cat	gtgg	catg	acaa	ccaaag	cttaac	ccac	gcctaa	gcgtgac	3660
QY	4175	aat	ggcaag	gcatgt	catgtac	gacag	ccaaaa	tggtca	aaata	ttaccac	4234

Qy	3704	ggtgtgttaggtgctggcattgatggccaaactctgattaccagagatgaattggcttt	3763
Db	2956	ggtaattcaagcactggcattgatggcaagcgtatccaccaagatcaaatggcttt	3015
Qy	3764	actggacataaaggctcacttgataaagacaaaacccacttaagcaaacg	3814
Db	3016	actggggctaaaggctcacttgataccacaaaacccacccacttaaccaaagacaagcttaaa	3075
Qy	3815	-----ggcattaacgcaggtgtaaaaaagattaccaac	3847
Db	3076	gtgggtgaagttgaaattaccaacactggcattaacgcaggtgtaaaaagattaccaac	3135
Qy	3848	attcaatcagggtgaagattggccaaaacacgacatgactgtgcagagcggcaagatttat	3907
Db	3136	attcaatcagggtgattaccacaaacagcaatgatgtgtgcagcggtcggtgtttat	3195
Qy	3908	gatttaaaaaacggaacttgaaaaacaaatcagcagtagctgccaaaaacagcacaaaaactca	3967
Db	3196	gatttaaaaaacggaacttgaaagcaaaatccacagtgctctaaaaacagcacaaaaactca	3255
Qy	3968	ttacacgaattctcagtagcagatgaacaaaggtaaataactttacggttagtaaccccttac	4027
Db	3256	ttacacgaattctcagtagcagatgaacaaaggttaactcatttcaggttagtaaccccttac	3315
Qy	4028	tcagttatgcacacctcaaaagacctctgatgtcatcactttgcaggtgaaaaacggcatt	4087
Db	3316	tcagttatgcacacctcaaaagacctctgatgtcatcactttgcaggtgaaaaacggcatt	3375
Qy	4088	accaccaaggtataaagggtgtgtcgctgggcatggacattaccacaaaccaaaggtctaac	4147
Db	3376	accaccaaggtataaagggtgtgtcgctgggcatggacattaccacaaaccaaaggtctaac	3435
Qy	4148	acgcctaaagctgacogtgggttaataatgccaagggcattgtcattgcagaccgcaaaat	4207
Db	3436	acgcctaaagctgacogtgggttaataatgccaagggcattgtcattgcagaccgcaaaat	3495
Qy	4208	ggtcaaaatcacatcacaggaactaagcaaacactctagcttaagttaaccaatgataaagg	4267
Db	3496	ggtcaaaatcacatcacaggaactaagcaaacactctagcttaagttaaccaatgataaagg	3549
Qy	4268	agcgtacgcacacagaaacagggcaataatacctaaagacgaagacaaaacccgtgcgcgc	4327
Db	3550	ggtgcaggaacgcactaagcgaagggttgccaatgacacggacaaaacccgtgcgcgc	3609
Qy	4328	agcattgttgatgtctaaagcgcaggtttaaacttgcaaggaactgtggaagcgtgtgac	4387
Db	3610	agcattgttgatgtctaaagcgcaggtttaaacttgcaaggaactgtggaagcgtgtgac	3669
Qy	4388	tttgtctcaattatgacaccgtcaactttgccatggcaatgccaccgcgctaagggtg	4447
Db	3670	tttgtctcaattatgacaccgtgtgactttatcgatggcaatgccaccgcgctaagggtg	3729
Qy	4448	acctatgatgcacaagcaaaacccagtaaaagtgtgtctatgatgtcgaattgtggaatga	4507
Db	3730	acctatgatgcacaagcaaaacccagtaaaagtgtgtctatgatgtcgaattgtggaatga	3789
Qy	4508	accattgaagttaaa--gataaaaaacttgggttaaaaaacccacattgacacagttact	4564
Db	3790	accattgaagttgacaagtgataaaaaacttgggtcctaaacccacacacattgacacaaa	3849
Qy	4565	ggcacaggctgctaataaatttgccctaaagcaatcaagctactgcgcgtgcgctgtgcaag	4624
Db	3850	agtgtctaattggttaatgca--accataatttagtgcgcgcggtgcggtatgccttgtttaa	3906
Qy	4625	ggcagtgatatcgttgcctatcaaacaccttatctgggcacatccaaactgccaaagg	4684
Db	3907	ggcagtgatatcgcaccacctataaataaccttggctggcgtgacatccaaacgcgcaagg	3966
Qy	4685	gcaagccaagcgaaacaactcgaagcgttatgtgatgctgtatggcaataaggtoacttat	4744
Db	3967	gcaagccaagcgaagcgtcgaagcgttatgtgatgctgtatggcaataaggtoacttat	4026

QY	4745	gacagtaccgataaacaag	tactatcatcaagccaaaat	tgatggcagctgataaaacaaa	4804
DB	4027	gacgttacccgataaagaag	tactatcaagtcaat	tgacaagggtcaagtggacaaaacaaa	4086
QY	4805	gaagttgccaaagacaaaac	tgctgcgccaagccaaaccccaagatggcacattggctcaa	4864	
DB	4087	gaagttgccaaagacaaaac	tgctgcgccaagccaaaccccaagatggcacattggctcaa	4146	
QY	4865	atgaattgtaaatcagctcat	taacaaagaacaaagtaaat	tgatgccataaaaaagcaaggc	4924
DB	4147	atgaattgtaaatcagctcat	taacaaagaacaaagtaaat	tgatgccataaaaaagcaaggc	4206
QY	4925	atcaatgaagacaaacgcct	tgttaaaggacttgaaaaagccgccttctgatacaaaaacc	4984	
DB	4207	atcaatgaagacaaacgcct	tgttaaaggacttgaaaaagccgccttctgatacaaaaacc	4266	
QY	4985	aaaaacgcgcagtaactgt	gggtgatttaaatgcggttgccaaacacgcgtgacacctt	5044	
DB	4267	aaaaacgcgcagtaactgt	gggtgatttaaatgcggttgccaaacacgcgtgacacctt	4326	
QY	5045	gcaggggatcacaggacaa	cgcgttaaaaaactggcgagactttgacctcaaaagtgggg	5104	
DB	4327	gcaggggatcacaggacaa	cgcgttaaaaaactggcgagactttgacctcaaaagtgggg	4386	
QY	5105	caaacagacaccaataa	gctaaccgataatacatcgggtgtgtagcagctactgatggc	5164	
DB	4387	caaacagacaccaataa	gctaaccgataatacatcgggtgtgtagcagctactgatggc	4446	
QY	5165	ttcactgtcaaaactgcca	aaagacctaacaatcttaacagcgttaatgcagtggtgcacc	5224	
DB	4447	ttcactgtcaaaactgcca	aaagacctaacaatcttaacagcgttaatgcagtggtgcacc	4506	
QY	5225	aaaaattgatcaaaaagc	gtctctttgtagactcaagcgggtcaagccaaagcaaacacc	5284	
DB	4507	agaattgatgaaaaaggg	catctctttgtagacgcaaaagcgttcaagccaaagcaaacacc	4566	
QY	5285	cctgtgctaagtgc	caattgggctggacctgggtgggc-----	5320	
DB	4567	cctgtgctaagtgc	caattgggctggacctgggtgggc-----	4626	
QY	5321	-----	-----	5320	
DB	4627	gctgttgatgataacg	atgctgggtgaactttaagcagtttaataagattgccaacacggtc	4686	
QY	5321	-----	-----	5320	
DB	4687	aaacacctaaacaa	ccaaagtaactcaggtgcgttaacccttgtgtgtaaccgataggc	4746	
QY	5321	-----	-----	5320	
DB	4747	aatggcaagcccatcaa	tgtggcaccgatggcaagcccccaaaagccatcaagggtccgat	4806	
QY	5321	-----	-----	5320	
DB	4807	ggtaatactatacgc	caacgcgaacgcgtacctgtggacaaagatggcaagccatc	4866	
QY	5321	-----	-----	5320	
DB	4867	accgatgcggacaaa	acctgccaaatcggcagctcatggcaaaccccttgatgcaggtcat	4926	
QY	5321	-----	-----	5320	
DB	4927	caagtggtggcaagc	ctagcggcgaactcagatggccatcaccttaaccaacatcaagtcc	4986	
QY	5321	-----	-----	5320	
DB	4987	actttgccacaaat	tgacacacaaacacaggtaaatggcgaatgcagggtgcaagcccaagt	5046	
QY	5321	-----	-----	5320	
DB	5047	ctgccagcctatcag	cagacagcaagtaattgctgcaggtgtcaagaatgtgctaaat	5106	
QY	5321	-----	-----	5320	

Db	6187	atcgggtgataacgcacaaagccacggcgatcattccatcgccatcggtacagagcaatgtg	6246
QY	5987	gtacgaggtaaacactctgggtgcacatcggcgacccaagcactgttaaggctgataacagt	6046
Db	6247	gtacaaggtaagcaactctgggtgccatcggcgacccaagcactgttaaggctgataacagt	6306
QY	6047	tacagtgtgggtataatacaaacccagtttaccgaatgccactcaaacccgatgtctttggtgtg	6106
Db	6307	tacagtgtgggtataatacaaacccagtttctgatgccactcagccgatgtctcttgggtgtg	6366
QY	6107	ggcaataaacctcacggtgaccgaagaactcgggttcgcttaggttcaaaactctgcact	6166
Db	6367	ggcaataaacctcacggtgaccgaagaactcgggttcgcttaggttcaaaactctgcact	6426
QY	6167	agtcgaggcacacacgacgagccacaaagcccaaaaaattgacgcacagcaggtacaacc	6226
Db	6427	agtcgaggcacacacgagggcacaaagcccaaaaaattgacgcacagcaggtacaacc	6486
QY	6227	accaacgaggtgcaaccggtacggtttaaaggcttctgtgacaaaacggcggttgggtgcg	6286
Db	6487	accaacgaggtgcaaccggtacggtttaaaggcttctgtgacaaaacggcggttgggtgcg	6546
QY	6287	gtctccggtgggtcctcaggtgtgaacgcctatccaaaaatgtggcagcaggttgaggtc	6346
Db	6547	gtctccggtgggtcctcaggtgtgaacgcctatccaaaaatgtggcagcaggttgaggtc	6606
QY	6347	agtgccacacgaaccatcggttcaatgttcacaggttcgtacaagccacccaaagcatt	6406
Db	6607	agtgccacacgaccatcggttcgaatgttcacaggttcgtacaagccacccaaagcatt	6666
QY	6407	gccaacgcaaccaatgagcttgaccatcttatccacaaaaacgaaaaataagggccaatgca	6466
Db	6667	gccaacgcaaccaatgagcttgaccatcttatccacaaaaacgaaaaataagggccaatgca	6726
QY	6467	gggatttcaatcagcgtatggcgttggcgtccatgccacaagcctacatccctggcagatcc	6526
Db	6727	gggatttcaatcagcgtatggcgttggcgtccatgccacaagcctacatccctggcagatcc	6786
QY	6527	atggtttaccgggggtattgccaccacaacggtcgaagggtggggtggcaggtgggactgtcg	6586
Db	6787	atggtttaccgggggtattgccaccacaacggtcgaagggtggggtggcaggtgggactgtcg	6846
QY	6587	aagctgtcggataatgctcaatgggtatttataaatcaatggttcagccgatacccaaggc	6646
Db	6847	aagctgtcggataatgctcaatgggtatttataaatcaatggttcagccgatacccaaggc	6906
QY	6647	catgtaggggcgcgagttgtgtcaggtttttcactt	6682
Db	6907	catgtaggggcgcgagttgtgtcaggtttttcactt	6942

RESOLUT TO
AAAF59129
ID AAAF59129 standard: DNA: 720 BP.

AC
XX
DT

XX Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW oritis media; detection; ds.
KW

XX WO200107619-A1

26-JUL-2000: 2000WO-CA00870.

XX
T 7

[illegible]

QY	6260	tttctggaacaaacgcgggttggtcggtctccgtggtgacctcagtgctgaacgcggt	6319
Db	2615		
Db	2615	tttgcggtgcaacgcgcacggttcgttcgttcggtcgcaagcgcgaagaagacgt	2674
QY	6320	atccaaaattggcagcaggtagtgcagtgccaccagcacgatcggtcaatggtagc	6379
Db	2675		
QY	6380	cagttgtacaaggccaccaaacgattgccaacgcaaccaatgacttgaccatcgtatc	6439
Db	2735		
QY	6440	caccaaacyaaaaataaggccaatgcagggtatttoatcagcgatggcgatggcgatc	6499
Db	2783	aataaagtggaacgctgcatgcaggtacagcaagtcattagcgtcttcacagtta	2842
QY	6500	ccacaagcctacattctggcagatccatggttacccggggtattgccaccacaacggt	6559
Db	2843		
QY	6560	caagtgctggtggcagtggaactgcgaagctgtcggataatggtcaatgggtatttaa	6619
Db	2903		
QY	6620	atcaatggttcagcgcatacccaggccatgtagggcgcgagtggtgtaggtttcac	6679
Db	2963		
QY	6680	ttttaa 6685	
Db	3023		
		tgttaa 3028	
 RESULT 14			
ID	AAA92495	standard; DNA; 3354 BP.	
XX	AAA92495;		
DT	17-JAN-2001	(first entry)	
DE	Haemophilus influenzae adhesin (Hia) gene from NTHI strain K9.		
KW	Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine		
KW	non-typeable Haemophilus influenzae; anti-inflammatory; auditory;		
KW	antibacterial; meningitis; epiglottitis; septicaemia; otitis media;		
KW	diagnosis; immunogenic; antigen; ds.		
OS	Haemophilus influenzae.		
PN	WO200055191-A2.		
PD	21-SEP-2000.		
PF	16-MAR-2000; 2000WO-CA00289.		
PR	16-MAR-1999; 99US-0268347.		
PA	(CONN-) CONNAUGHT LAB LTD.		
PI	Loosmore SM, Yang Y, Klein MH;		
WP	WPI: 2000-618897/59.		
DR	P-PSDB; AAE23858.		
PT	Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for		
PT	use as antigens and vaccines and for treating Hemophilus influenzae		
PT	infection		
PS	Claim 1; Fig 22; 275pp; English.		
CC	The present sequence represents a Haemophilus influenzae adhesin (Hia)		

Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Hemophilus influenzae infection -
 Claim 1; Fig 22; 275pp; English.
 The present sequence represents a Haemophilus influenzae adhesin (Hia)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:39:39 ; Search time 28653.2 Seconds
(without alignments)
3595.553 Million cell updates/sec

Title: US-09-361-619-5

Perfect score: 6972

Sequence: 1 ccattgatattggcagggtgt.....ccgtgacctgcaggtcgac 6972

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_NA_Main.*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US098_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US099_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6972	100.0	6972	17	US-09-361-619-5
2	6961	99.8	6973	8	US-08-483-855-1
3	6961	99.8	6973	10	US-08-621-944-1
4	6961	99.8	6973	10	US-08-621-944A-1
5	6961	99.8	6973	13	US-08-945-567-1
6	6961	99.8	6973	13	US-08-945-567A-1
7	6961	99.8	6973	13	US-08-945-567C-1
8	6961	99.8	6973	13	US-08-945-567D-1
9	6961	99.8	6973	17	US-09-361-619-1
10	6939	99.5	6975	8	US-08-431-718-1
11	6939	99.5	6975	8	US-08-431-718A-1
12	6939	99.5	6975	8	US-08-431-718C-1
13	6258	89.8	6259	17	US-09-361-619-12
14	6141	88.1	6141	17	US-09-361-619-6
15	5976	85.7	5976	10	US-08-621-944-2
16	5976	85.7	5976	10	US-08-621-944A-2
17	5976	85.7	5976	13	US-08-945-567-2
18	5976	85.7	5976	13	US-08-945-567A-2
19	5976	85.7	5976	13	US-08-945-567C-2
20	5976	85.7	5976	13	US-08-945-567D-2
21	5976	85.7	5976	17	US-09-361-619-2
22	4671.2	67.0	62909	22	US-09-596-002-32
23	4671.2	67.0	62909	46	US-60-140-121-32
24	4655.8	66.8	15958	38	US-60-068-139-381
25	4655.8	66.8	17368	40	US-60-082-116-290
26	4396.8	63.1	9542	31	US-09-813-214-9
27	4101.6	58.8	6159	17	US-09-361-619-8
28	2590	37.1	6942	17	US-09-361-619-10
29	2007.2	28.8	6429	21	US-09-540-236-1539
30	2007.2	28.8	6429	44	US-60-128-476-1992
31	220.8	3.2	1005	21	US-09-540-236-1266
32	220.8	3.2	1005	44	US-60-128-476-2460
33	148	2.1	3030	16	US-09-268-347-29
34	146.4	2.1	3036	16	US-09-268-347-23
35	121.4	1.7	123	44	US-60-128-476-1993
36	85.4	1.2	7486	14	US-09-077-098-5
37	85.4	1.2	7486	14	US-09-077-098A-5
38	84.4	1.2	6091	18	US-09-453-702B-200
39	83.8	1.2	2079	16	US-09-268-347-25
40	83.8	1.2	8930	14	US-09-077-098A-1
41	83.8	1.2	8930	14	US-09-077-098-1
42	82.8	1.1	3294	16	US-09-268-347-43
43	78.4	1.1	1812	16	US-09-268-347-37
44	78.2	1.1	3300	16	US-09-268-347-31
45	78.2	1.1	7253	16	US-09-268-347-35

QY 1741 aaaagaggctgataataagtgctgaaagttaaacttgcataaaactttaaacatttaacatctttac 1800
|||||
Db 1741 aaaagaggctgataataagtgctgaaagttaaacttgcataaaactttaaacatctttac 1800
|||||
QY 1801 tgaggtgaatacaactacattaaatgcacaaacacacagtttaagtagtagtagtagtag 1860
|||||
Db 1801 tgaggtgaatacaactacattaaatgcacaaacacacagtttaagtagtagtagtagtag 1860
|||||
QY 1861 tactacagtgtaattattagtgtagtagtttaacctttaccagcccaatacagcgagtc 1920
|||||
Db 1861 tactacagtgtaattattagtgtagtagtttaacctttaccagcccaatacagcgagtc 1920
|||||
QY 1921 aagcacaaagcaaaacccgtctatgcgttaattggggtgaagtttactaataatgcagaaac 1980
|||||
Db 1921 aagcacaaagcaaaacccgtctatgcgttaattggggtgaagtttactaataatgcagaaac 1980
|||||
QY 1981 aacagcagcaatcgccactactcgtattaccagagataaaattggcttgcgcagatgg 2040
|||||
Db 1981 aacagcagcaatcgccactactcgtattaccagagataaaattggcttgcgcagatgg 2040
|||||
QY 2041 tgatgttgatgaaaaaacaagcaccattttggataaaaaaacaacttaagtggttagtgt 2100
|||||
Db 2041 tgatgttgatgaaaaaacaagcaccattttggataaaaaaacaacttaagtggttagtgt 2100
|||||
QY 2101 tgcaattaccatagacaatggcatttgatgcaggttaataaaaaagatcagtaacttgc 2160
|||||
Db 2101 tgcaattaccatagacaatggcatttgatgcaggttaataaaaaagatcagtaacttgc 2160
|||||
QY 2161 aggtagcagtgctaacgatcggtgtaccatcgaaacagctcaaaagcccgcaagcctacttt 2220
|||||
Db 2161 aggtagcagtgctaacgatcggtgtaccatcgaaacagctcaaaagcccgcaagcctacttt 2220
|||||
QY 2221 aaacgagggcgctggcatcagtgctcacacctactgaaatcacgttgatgtaagagtg 2280
|||||
Db 2221 aaacgagggcgctggcatcagtgctcacacctactgaaatcacgttgatgtaagagtg 2280
|||||
QY 2281 caatgttacgcccccacttacacattggcgtgaaacaccagcgcttaacagtgatgg 2340
|||||
Db 2281 caatgttacgcccccacttacacattggcgtgaaacaccagcgcttaacagtgatgg 2340
|||||
QY 2341 cactagtgataaatttagtggttaagggtagtggtacgaaacaataagcttagttaccgcga 2400
|||||
Db 2341 cactagtgataaatttagtggttaagggtagtggtacgaaacaataagcttagttaccgcga 2400
|||||
QY 2401 acatttggcaagctatctaaatgaagtcacacccctactgaaatcacgttgatgtaagagtg 2460
|||||
Db 2401 acatttggcaagctatctaaatgaagtcacacccctactgaaatcacgttgatgtaagagtg 2460
|||||
QY 2461 taccgttaaaagaagaacgatgatgcgcgaacgcgtatcacctggcgttaagatagcag 2520
|||||
Db 2461 taccgttaaaagaagaacgatgatgcgcgaacgcgtatcacctggcgttaagatagcag 2520
|||||
QY 2521 aaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctaaacggttgc 2580
|||||
Db 2521 aaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctaaacggttgc 2580
|||||
QY 2581 taccataaaagatggttacggttacctttggccttagccaagatagcggtctgaccattgg 2640
|||||
Db 2581 taccataaaagatggttacggttacctttggccttagccaagatagcggtctgaccattgg 2640
|||||
QY 2641 caaaagcaccctaaacaacgatgcttgactgttaagatataccaacgaaacaaatccaaagt 2700
|||||
Db 2641 caaaagcaccctaaacaacgatgcttgactgttaagatataccaacgaaacaaatccaaagt 2700
|||||
QY 2701 cgggtgtaattggcattaaattactaatgtgaaatgggtagtaatccaggtactggcattgc 2760
|||||
Db 2701 cgggtgtaattggcattaaattactaatgtgaaatgggtagtaatccaggtactggcattgc 2760
|||||
QY 2761 aaataccgcgtcattaccagagataaaattggcttgcgttgcgttgcgtgagtgagtgga 2820
|||||
Db 2761 aaataccgcgtcattaccagagataaaattggcttgcgttgcgttgcgtgagtgagtgga 2820
|||||
QY 2821 tacaacaaaccccttatcttgatcaagacaagctacaagttggcgaatgtaagattacca 2880
|||||

Db 2821 tacaacaaaccccttatcttgatcaagacaagctacaagttggcaatgtaagattacca 2880
|||||
QY 2881 cactggcattaacgcaggtggttaaagccatcacagggtgtcccaacacacgtcctagcat 2940
|||||
Db 2881 cactggcattaacgcaggtggttaaagccatcacagggtgtcccaacacacgtcctagcat 2940
|||||
QY 2941 tgcgcatacaaaagtagcgcgaacatagaaactgggcaatacaatccaaagacaagacaatc 3000
|||||
Db 2941 tgcgcatacaaaagtagcgcgaacatagaaactgggcaatacaatccaaagacaagacaatc 3000
|||||
QY 3001 caacgctgcagcatttaattgatatattaaatcacaggcttttaacctaaaaataataacaa 3060
|||||
Db 3001 caacgctgcagcatttaattgatatattaaatcacaggcttttaacctaaaaataataacaa 3060
|||||
QY 3061 cccattgactttgtctccacttatgacattgttgactttgccaatggcaatgccaccac 3120
|||||
Db 3061 cccattgactttgtctccacttatgacattgttgactttgccaatggcaatgccaccac 3120
|||||
QY 3121 cgccacagtaaccctgataccgctacaacaaacagtaaaagtggtatagtggaatgt 3180
|||||
Db 3121 cgccacagtaaccctgataccgctacaacaaacagtaaaagtggtatagtggaatgt 3180
|||||
QY 3181 ggaatgatacaaccattcatctaacaggcactgagacaataaaaaacttggcgtcaaac 3240
|||||
Db 3181 ggaatgatacaaccattcatctaacaggcactgagacaataaaaaacttggcgtcaaac 3240
|||||
QY 3241 caccaaactgaacaaaaacaagtgcttaattggttaatacagacaacttaatttaactc 3300
|||||
Db 3241 caccaaactgaacaaaaacaagtgcttaattggttaatacagacaacttaatttaactc 3300
|||||
QY 3301 tagtgagaagtgcctttgttaacgccaagacatcgccgaaaaatcaaacacccctagc 3360
|||||
Db 3301 tagtgagaagtgcctttgttaacgccaagacatcgccgaaaaatcaaacacccctagc 3360
|||||
QY 3361 caaagaaattcacaccacccaaaggcacagcagacacccctacaaaaactttaccgttaa 3420
|||||
Db 3361 caaagaaattcacaccacccaaaggcacagcagacacccctacaaaaactttaccgttaa 3420
|||||
QY 3421 aaaggtagatgaaaaataataatgctgtagcgcaaacccatcacctgggtgcaaaagaa 3480
|||||
Db 3421 aaaggtagatgaaaaataataatgctgtagcgcaaacccatcacctgggtgcaaaagaa 3480
|||||
QY 3481 cgcaataatacaagtcacacccctaacactcaaaaggtgaaacggtcttaattataaac 3540
|||||
Db 3481 cgcaataatacaagtcacacccctaacactcaaaaggtgaaacggtcttaattataaac 3540
|||||
QY 3541 cgcaaaaaatggttacggttacctttggcatttaacacacacagcggtcttaaaagccggcaa 3600
|||||
Db 3541 cgcaaaaaatggttacggttacctttggcatttaacacacacagcggtcttaaaagccggcaa 3600
|||||
QY 3601 aagcacccctaaacgagcggtggttgccttatataaaacccctggttagcgacaatacca 3660
|||||
Db 3601 aagcacccctaaacgagcggtggttgccttatataaaacccctggttagcgacaatacca 3660
|||||
QY 3661 agtcggtgctgattggcgtagagtttgcgaaggttaataatgggtgttaggtgctgg 3720
|||||
Db 3661 agtcggtgctgattggcgtagagtttgcgaaggttaataatgggtgttaggtgctgg 3720
|||||
QY 3721 cattgtaggcacaactcgattaccagagatgaaattggcttactgggaactaatggctc 3780
|||||
Db 3721 cattgtaggcacaactcgattaccagagatgaaattggcttactgggaactaatggctc 3780
|||||
QY 3781 acttgataaaagcgaaccccccttaagcaaaagcggttaacgcaggtggttaaaagat 3840
|||||
Db 3781 acttgataaaagcgaaccccccttaagcaaaagcggttaacgcaggtggttaaaagat 3840
|||||
QY 3841 taccacaattcaatcaggtgagattgcccacaaacagccatgatcgtgacagcgcgcaa 3900
|||||
Db 3841 taccacaattcaatcaggtgagattgcccacaaacagccatgatcgtgacagcgcgcaa 3900
|||||
QY 3901 gatttatgattataaaacccgaacttgaaaaacaaatcagcgactgctgcacaaacagcaca 3960
|||||

Db	3901	gattttatgatttaaaaccggaacttgaaacaaaatacagcag tactgcgcaaaacgacaca	3960
Qy	3961	aaactcatcacaggaattctcagtagcagatgaacaaggtaataaactttacggttagtaa	4020
Db	3961	aaactcattacacgaattctcagtagcagatgaacaaggtaataaactttacggttagtaa	4020
Qy	4021	cccttactccagttatgacaaactcaagacccctctgattgctcatcacctttgcagggtgaaa	4080
Db	4021	cccttactccagttatgacaccccaagacccctctgattgctcatcacctttgcagggtgaaa	4080
Qy	4081	cgcattaccaccaaggtaataaagggtggtgctggtggtgcatgacacaaaccaaagg	4140
Db	4081	cgcattaccaccaaggtaaaataaagggtggtgctggtggtgcatgacacaaaccaaagg	4140
Qy	4141	cttaaacacgcttaagctgacgcgtgggttaataataatggcaaggcattgtcattgacag	4200
Db	4141	cttaaacacgcttaagctgacgcgtgggttaataataatggcaaggcattgtcattgacag	4200
Qy	4201	ccaaaatggtcaaaatacaatacagaggactaagcaaacactctagctlaaagttaccaatga	4260
Db	4201	ccaaaatggtcaaaatacaccatacaggactaagcaaacactctagctlaaagttaccaatga	4260
Qy	4261	taaaagtgagctgacgacacagaacagggaataataatcaaaagacgaagacaaaaccg	4320
Db	4261	taaaagtgagctgacgacacagaacagggaataataatcaaaagacgaagacaaaaccg	4320
Qy	4321	tgcgcacgacttggttgattgactaagcgagcgctttaacttgcgaagcaatggtgaacg	4380
Db	4321	tgcgcacgacttggttgattgactaagcgagcgctttaacttgcgaagcaatggtgaacg	4380
Qy	4381	ggttgactttgctccacttatgacacgctcaacttggcgatggcaatgccaccaccgc	4440
Db	4381	ggttgactttgctccacttatgacacgctcaacttggcgatggcaatgccaccaccgc	4440
Qy	4441	taagtgtagcctatgatgacacagaacaaaccagtaaaagtggtctatgatgtcaattgga	4500
Db	4441	taagtgtagcctatgatgacacagaacaaaccagtaaaagtggtctatgatgtcaattgga	4500
Qy	4501	tgatcacaccattgaagttaaagataaaaaaacttggcgtaaaaaaccaccacattgaccag	4560
Db	4501	tgatcacaccattgaagttaaagataaaaaaacttggcgtaaaaaaccaccacattgaccag	4560
Qy	4561	tactggcacaggtgcttaataaaattgcccctaagcaatacgaactctctggcgatgcgttgt	4620
Db	4561	tactggcacaggtgcttaataaaattgcccctaagcaatacgaactctctggcgatgcgttgt	4620
Qy	4621	caaggcagtgatatgctgctcatcaaacaccttatctggcgacatccaaactgcca	4680
Db	4621	caaggcagtgatatgctgctcatcaaacaccttatctggcgacatccaaactgcca	4680
Qy	4681	aggggcaagccaagcgaacaaactcagcaggctatgtggatgctgatggcaataaggtcat	4740
Db	4681	aggggcaagccaagcgaacaaactcagcaggctatgtggatgctgatggcaataaggtcat	4740
Qy	4741	ctatgacagtgaccgatataacaagtaactatcaagccaaaataatgatggcacagtgtataaac	4800
Db	4741	ctatgacagtgaccgatataacaagtaactatcaagccaaaataatgatggcacagtgtataaac	4800
Qy	4801	caagaagttgccaaaagacaaaactggtgcccaagcccaaacccagatggcacattggc	4860
Db	4801	caagaagttgccaaaagacaaaactggtgcccaagcccaaacccagatggcacattggc	4860
Qy	4861	toaaatgaattcgaatacagttcattacaagaagacaaagtaaatgatgccataaaaaagca	4920
Db	4861	toaaatgaattcgaatacagttcattacaagaagacaaagtaaatgatgccataaaaaagca	4920
Qy	4921	agggcatcaatgaagacaaacgctttgttaaaaggacttgaaaaagcgcgttctctgataacaa	4980
Db	4921	agggcatcaatgaagacaaacgctttgttaaaaggacttgaaaaagcgcgttctctgataacaa	4980
Qy	4981	aacccaaaacgcgcagtaactgtggtgtattbaatgcgcttgcccaaaaacgcgctgac	5040
Db	4981	aacccaaaacgcgcagtaactgtggtgtattbaatgcgcttgcccaaaaacgcgctgac	5040


```

GENERAL INFORMATION:
APPLICANT: SASAKI, Ken
APPLICANT: HARKNESS, Robin E.
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: KLEIN, Michael H.
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
MEMBRANE PROTEIN OF MORAXELLA
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/621,944
FILING DATE: 26-MAR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/478,370
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-587
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 708..6683
PS-08-621-944-1

```

Query Match	99.8%;	Score 6961;	DB 10;	Length 6973;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 6972;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps
QY 1	ccatgcatatgggcaagtgctgcgcctgcccgtatgatggcgatgacaccccatittgcc	60		
Db 1	CCATGATGATGGCAGGTGTGCGCTGCCGTATGATGGCATGACCCCATTTGCC	60		
QY 61	catatctgtacatgttgacatgtgatgatattaacatgtgacatgattaacatgttt	120		
Db 61	CATATCTGTGACGATTTGACATGTGATATGATTTACATCTGCACATGATTTACATTTGTT	120		
QY 121	aatactgttgccatcattaccataatttgtaacgcatttagtaacgcatttggtaaaaat	180		
Db 121	AATACTGTTGCCATCATTTACCAATAATTTAGTAACGCATTTAGTAACGCATTTGTAATAAT	180		
QY 181	catggcgcacctttatgtgtatcatatgaatagaatattatgatgtatctctgatttgt	240		
Db 181	CATTGGCCCTTTATGTGTATCATATGAATAGAATATTTATGATTGTATCTCATTTTGT	240		
QY 241	atcagaatggtgaatgcctataataatgacctacgagttgatttgggttaataatcacctctatg	300		
Db 241	ATCAGAAATGGTGATGCTATATGATGATGCTACGAGTTGATTGGGTAAATCACCTCTATG	300		
QY 301	atttgatataattttgaaacatacttattgaacttaaacaccataatggttataatttagca	360		

Db	301	ATTTGATATATTTTGAACATAATCTTATGACATTAAATCACCATATGGTTATATAATTAGCA	360
Qy	361	taatggttagcttttggtaaaatcacatcgcaatattgttctactgttactaccatgct	420
Db	361	TAAATGGTAGGCTTTTGTAAAAATCACATCGCAATATGTTCTACTGTTTACTTACCATGCT	420
Qy	421	tgaatgacgatcccaatcacacagattcattcaagtatgttggttatcagcaccatttta	480
Db	421	TGAATGACGATCCCAATCACACGATTCATTCAGTGTGTTGTFATACGCCACATTTA	480
Qy	481	ccctaataattcaatcaaatgacctatgtctgaatgtatcatctttttaaggtaaacac	540
Db	481	CCCTTAATATTTCATCAAAATGCTATGTACGATGTATCATTTTTTTTAAAGGTAACAC	540
Qy	541	catgaatcacatcatataaagtcatctttaacaaagccacagcacatttatggcagtcg	600
Db	541	CATGAATCACATCTATAAGTCATCTTTTACAAAGCCACAGGCATTTATGGCAGTGGC	600
Qy	601	agagtacgccaaaatcccaacagcac-ggggggggtagctgtgtctacaggaagtggca	659
Db	601	AGAGTACGCCAAATCCACACAGCACGGGGGGGTAGCTGTCTACAGGCCANGTTGCA	660
Qy	660	gtgtatgcactctgagctttgcccgttatgtccgcgtctgcgtctctctgtatcggtgcaa	719
Db	661	GTGTATGACACTCTAGCTTTTGCOCGTATTTCGCCGCTCGCTCTCGTGTATCGGTGCAA	720
Qy	720	cgttcagtgcagtgcttatgtctcaaaaaaagataccaacaatcgcgaattggtgaac	779
Db	721	CGCTACGTGGCAGTGCTTATGTCTCAAAAAAAGATACCAAAATATPCGCAATTTGGTGAA	780
Qy	780	aaaaccgccaadagcgtcagcgcactgcgaagcgagcggtgatcgagccattgctattg	839
Db	781	AAAACCAACCAAGACGCTCAGGCACATGCCAACGGCGACGGTGATCGAGCCATTGTATTG	840
Qy	840	gtgaaaatgctaacgcacagggcggttcaagccatcgccatcggttagtagtaataaaactg	899
Db	841	GTGAAAATGTCTAACGCACAGCGGGGTCAAGCCATCGCCATCGGTAGTACTTAATAAACTG	900
Qy	900	tcaatggaagcagtttggataaagatagggtacagatgctacaggggtcaagaagtcctatcgcca	959
Db	901	TCAATGGAAGCAGTTTGGATTAAGATAGGTACCCGATGCTACGGGTCAAGAGTTCATTCGGCA	960
Qy	960	tcggtggtgatgataaaggctagtggtgatgctcgatgccatcggttagtgatgacttac	1019
Db	961	TCGGTGGTGATGTAAAGGCTAGTGGTGATGCCCTCGATTCGCCATCGGTAGTGATGACTTAC	1020
Qy	1020	atttgcttgatcagcatggttaatctctaaacatccgaaggtactctgtattaacgatctta	1079
Db	1021	ATTTGCTTGATCAGCATGGTAATCTTAACATCCGAAGGTACTCTGATTAACGATCTTA	1080
Qy	1080	ttaacggccatcgagttataaagaataacgaagctcaagaagataatgatgtaaatatata	1139
Db	1081	TTAAACGCCCATGTCAGTATTAAAAAGAAATACGAAGCTCAAAAGGATTAATGATGTAAATATA	1140
Qy	1140	gacgcacaaacgaagcgagacgcagttactgcagtgagggccatgtcatatgcacagg	1199
Db	1141	GACGCACACCGCANGCGGACAGCCAGTACTGTCAGTGGGCGCATGTCTATATGCACAGG	1200
Qy	1200	gtcatattttccaacgcctttggttacacgggcaacagctaaagtgcttattccttggcag	1259
Db	1201	GTCAATTTTCCACGCCTTTTGGTACACGGGCAACAGCTAAAGTGCCATTATCTCTTGGCAG	1260
Qy	1260	tgggtcttgccgcacaaagccagggccaatctacaatcgctattggttctgtagcaacat	1319
Db	1261	TGGGTCTTTGGCGGCACAGCGAGGGCCAATCTACAAATCGCTATTGGTTCGTGATGCAACAT	1320
Qy	1320	ctagctcgttggagcagatagcccttgggtgcaggtactcgtctcagctacagggcagta	1379
Db	1321	CTAGCTCGTTGGGAGCGATATGSCCTTTGGTGCAGGTACTCGTCTCAGCTTACAGGCGCATTA	1380
Qy	1380	ttagcccttaggttcaaggttctgttgcactcagagtgataataatctcagaccggcctata	1439

Db 1381 TTGCCCTAGGTCAGAGTCTGTGTGTCACCTCAGAGTGATATAATTTCTAGACGGCGCTATA 1440
QY 1440 caccataatcaccagcagcactagaccccaagtttcaagccaccaataataatcgaagcggttc 1499
Db 1441 CACCAAAATACCCAGGACATAGACCCCAAGTTTCAAGCCACCAATAATAGAAAGCGGGTC 1500
QY 1500 cacttccatttgtagtaactctatcaaacgtaaaatcatcaatgctgggtaggtgtta 1559
Db 1501 CACTTTCATTTGTTAGTAACTCTATCAACAGTAAATCATCATCTGCGGTGAGGTGTA 1560
QY 1560 ataaacccgatcggttcaatgtggcacagctagaagcggtgtgtaagtggtgtaaggagc 1619
Db 1561 ATAAACCGATGCGGTCAATGTGGCACAGCTAGAAAGCGGTGTAAGTGGCTAAGGAGC 1620
QY 1620 gtagaattactttcaggtgtatgataacagctactcaactcaaaaatagtttggataata 1679
Db 1621 GTAGAATTTACTTTTCAGGGTGATGATAACAGTACTGACGTAAATAATAGTTTGGATAATA 1680
QY 1680 cttaactattaaaggtgggtgcagagaccaacgcattaaaccgataataatcgggtgtg 1739
Db 1681 CTTTAACTATTAAAGTGGTGACAGACCAACGCATTAAACCGATATAATATCGGTGTGG 1740
QY 1740 taaagaggtgtataatagttggtctgaaagttaaactgtcctaaactttaaacaacttta 1799
Db 1741 TAAAGAGGCTGATAATAGTGGTCTGAAAGTTTAAACTTGGTAAACCTTTAAACAATCTTA 1800
QY 1800 ctgaggtgaatacaactacattaaatgacacacacacagctaaaggttaggttagtagta 1859
Db 1801 CTGAGGTGAATACAACTPACATTTAAATGCCACACACAGTTAAAGTTAGTTAGTAA 1860
QY 1860 gtactacagctgaattattgagtgatgttttaacctttaccagcccaataacagggcagtc 1919
Db 1861 GTACTACAGCTGAATTTATGAGTGATAGTTTAACTTTTACCAGCCCAATACAGGCAGTC 1920
QY 1920 aaagcacaagcaaaacogtctatggtcgttaagtgggtgaagtttacttaataatgcagaaa 1979
Db 1921 AAMGCAAGCAAAACCGTCTATGGCGCTTAAATGGGTGAAGTTTACTTAATTAATGCAGAAA 1980
QY 1980 caacagcagcaactcggcactactcgtattaccagagataaaattggcttggctcagatg 2039
Db 1981 CAACAGCAGCAATCGGCACCTACTCGTATTACCAGAGATAAAATTTGGCTTTGCTCGAGATG 2040
QY 2040 gtgattgtgataaaacaagcaccatatttggataaaacaacttaaggtggtagtg 2099
Db 2041 GTGATGTTGATGATAAAACAAGCACCATTTTGGATAAAACAACCTTAAAGTGGTAGT 2100
QY 2100 ttgcaattaccatagacaaatggtcattgatgcaggtataaaagatcagtaaatcttgcca 2159
Db 2101 TTGCNAATTACCATAGACNATGGCATTTGATCGAGTATAAAAGATCAGTAATCTTGCCA 2160
QY 2160 aagtagcagtgtaacagatcggtttaccatcgaaacgctcaaaagccgaagcctactt 2219
Db 2161 AAGGTAGCAGTGTCAACGATCGGTTTACCATCGAACACAGCTCAAAAGCCGCCAAGCCTACTT 2220
QY 2220 taaacgagcgctlgcactcagtcactcactcagctgaatcagtttgatgctaaagtg 2279
Db 2221 TAAAGCAGGCGCTGGCATCAGTGTACACCCCTACTGAAATATCASTTGTATGCTAAGAGTG 2280
QY 2280 gcaattggtaccgcccacttaacacattggcgtgaaacaccacccagcgtttaacagtgatg 2339
Db 2281 GCAATGTTACCGCCCAACTTACAACATTTGGCGTGAAACCCACCGAGCTTAAACAGTGATG 2340
QY 2340 gcactagtgataaatttaggttgaaggttagtgtaagaaacaataagcttagttaccgccc 2399
Db 2341 CCACATGATCATAAATTTTATGTTAAGGGTAGTGGTACCAACAAATAGCTTACTTACCGCCG 2400
QY 2400 aacatttggcagctctatcaatgaagtcaatcgaaacggctgacagtgctctcaaaagct 2459
Db 2401 AACATTTGGCAGCTATCTAATGAAGTCAATCGAACGGCTGACAGTGCTCTCAAAAGCT 2460
QY 2460 ttaccgttaaagaagaagacgatgtagcgcacacgctatcacggtggtctaaagatacga 2519
Db 2461 TTACCGTTTAAAGAGAAGACGATGATGACGCCACGCTATCACCGTGGCTTAAAGATACGA 2520

QY 2520 caaaaaatgcggcgagctcagcatcttaaaactcaaaaggtataaaacgggtctaaacgggttg 2579
Db 2521 CAAAAAATGCCGGCGCAGTCAGCATCTTAAAACTCAAAAGGTAAAAAGCGGTCTAACGGTTG 2580
QY 2580 ctacccaaaaaagataggtaacggtttacccttgggcttagccaagatagcgggtctgaccattg 2639
Db 2581 CTACCCAAAAAAGATGGTACGGTTACCTTTGGGCTTTAGCCAAAGTAGCGGTCTGACCATTTG 2640
QY 2640 gcaaaagcaccctaaacaaacgctggcttgactgtttaagataaccaagaacaaatcccaag 2699
Db 2641 GCANAAGCACCCCTAAACAACGATGGCTTGACTGTTAAAGATACCAACGAACAAATCCAAAG 2700
QY 2700 tgggtgctaaatgggaatttaattacttaattgaatggtagtgaatcccaaggtactggcattg 2759
Db 2701 TCGGTGCTTAATGGCAATTAATTTACTAATGTAATGGTAGTAATCCAGGTACTGGCATTTG 2760
QY 2760 caaataccgcctgcgactaccagagataaaattggcttggctgggttctgattggtgagttg 2819
Db 2761 CAAATACCGCTCGCATTTACAGAGATAAATTTGGCTTTGCTGGTCTGATGGTGCAGTTG 2820
QY 2820 atacaaaaaaaccttatctctgtatcaagacaagctacaagttggcaattgtaagaataacca 2879
Db 2821 ATACAAACAAACCTTATCTTGATCAAGACAAGCTACAAGTTGGCAATGTTAAGATTACCA 2880
QY 2880 aactggcatttaacgacggtggtaaaagccatcaacagggctgtcccaacacactgcctagca 2939
Db 2881 ACACCTGGCATTTAACGACGAGTGGTAAAGCCATCACAGGGCTGTCCCAACACTGCGCTAGCA 2940
QY 2940 ttggcgaataaagtagcgcgcaacatagaaactgggcaatatacaatccaaagacaaagacaat 2999
Db 2941 TTGCGCATCAAAAGTAGCGCCACATAGAACTGGCAATACATCCAAAGACAAGACAAAT 3000
QY 3000 ccaacgctgccagcatttaattataattataacacagggctttaaacttaaaaaataaaca 3059
Db 3001 CCACGCTGCCAGCATTAATGATATATTAAATACAGGCTTTAACCTTAAAAATAATAACA 3060
QY 3060 acccattgaacttgcctcacttatgacattgttgactttgccaatggcaatggcaccaca 3119
Db 3061 ACCCAATGACTTTGTCTCCACTTATGACATTTGTTGACTTTGCCAATGGCAATGCCACCA 3120
QY 3120 cggccacagtaacccatgataccgttaacaaacacagtaaatgggtatgattgattgattg 3179
Db 3121 CGCCACAGTAACCCATGATACCGTTAACAAACACCAAGTAAAGTGTATATGATGTGAATG 3180
QY 3180 tggatgatacaaccattcttaacagcactgacatgaataaaaaacttggcggtcaaaa 3239
Db 3181 TGGATGATCAACCATTTCTATCAACAGGCACGTATGCAATTAANAACCTTGGCGTCANAA 3240
QY 3240 ccacccaaactgaacaaaaaagtgctaatggtaatacacagcaactaaactttaattgtaact 3299
Db 3241 CCACCAACTGAACAAAAACAAGTGTAAATGGTAAATACAGCACTAACTTTAATGTTAACT 3300
QY 3300 ctagtgtgaagatgccccttggtaacgcaaaagacatcgccgaaaaatctaaacacccctag 3359
Db 3301 CTAGTGATGAAGATGGCCTTGTTTAACGCCCAAGACATCGCCGAAAAATCTAAACCCCTAG 3360
QY 3360 ccaagaaatttcacacaccccaaaagggcacagcagacacccctcacaaaactttaccgtta 3419
Db 3361 CCAAGGAATTTACACCCACCCAAAGGCACAGCACCGCCCTTACAAACCTTTACCGTTA 3420
QY 3420 aaaaaggtagatgaaaaataatgtctgtagcgcacacgccaacgccaatccagtggggtcaaaaga 3479
Db 3421 AAAAGGTAGATGAAAAATAATGCTGATCACGCCCAACGCATCACCGTGGGTCAAAAGA 3480
QY 3480 acgcaataatacaagtaacacccctcaactcaaggtgaaacgggtcttaataataaaa 3539
Db 3481 ACGCAAAATATCAAGTCAACCCCTTACACTCAAAAGGTGAACGGGTCTTTAATATTAAAA 3540
QY 3540 ccgacaaaaatgggtacgggttaccttggcattaaacacacaaaggggtctttaagcgcgga 3599
Db 3541 CCGACAAAATGGTACGGTTACCTTTACCTTTGGCATTAACACCAACAGCGGTCTTTAAAGCGCGCA 3600

QY 3600 aaagcacccctaaacgacgggtggtctgtctctataaaacccccactggtgagcaacaaatcc 3659
DB 3601 AAAGCACCTTAACGACGGTGGCTGTCTATTAAAAACCCCACTGGTAGCAACAAATCC 3660
QY 3660 aagtcgggtgctgagtgagctgaagtttgcgaaggttaataataaagtggttgaagtgctg 3719
DB 3661 AAGTCGGTGTGATGGCGTGGAAGTTTGCCAGAGTTTAATTAATGAATGTTGAGTGCTG 3720
QY 3720 gcattgatggcacactgcattaccagagatgaaattggctttactgggactaatggct 3779
DB 3721 GCATTGATGGCACAACTCGCATTAACAGAGATGAAATTTGGCTTTACTGGGACTAATGGCT 3780
QY 3780 cacttgataaagcaaaccccccaactaagcaaacgagcattaaacgagtggttaaaaaaga 3839
DB 3781 CACTTGATAAAGCAAAACCCCACTTAAGCAAGACGGCATTAACGCAAGTGGTAAAAAGA 3840
QY 3840 ttaccaacattcaatcaggtgagatgtcccaaacacagcattgctgtgacagcgagca 3899
DB 3841 TTACCAACATTCAATCAGTGTGAGATTGCCCAAAACAGCCATGATGCTGTGACAGCGGCA 3900
QY 3900 agatttatgatttaaaacccgaacttgaaaacaaaatcagcagtlactgccaaaacagcac 3959
DB 3901 AGATTTATGATTAAAAACCGAACTTGAAACAAAAATCAGCAGTACTGCCAAAACAGCAC 3960
QY 3960 aaactcattcacagaattctcagtagcagatgaacaggtaataacttttcaggttagta 4019
DB 3961 AAAACTATTACAGCAATTCCTCAGTAGCAGATGAACAAGTAATTAACCTTTACGGTTAGTA 4020
QY 4020 acccttactcagttatgacacctcaagacctctgctcatcacctttcagagtgaaa 4079
DB 4021 ACCCTTACTCCAGTTATGACACCTTCAAGACCTCTGATGTATCACCTTTGCAGGTGAAT 4080
QY 4080 acggcattaccaccaaggttaataaagggtggtgctggtggcattgacccaaacaaag 4139
DB 4081 ACGGCATTACCACCAAGGTAAATAAAGGTGTGGTGGCGATTGACCAACCAAG 4140
QY 4140 gcttaacacagccttaagctgagtggtggtggttaataataatggcaaggcattgtcattgaca 4199
DB 4141 GCTTAACACAGCCCTTAAGCTGACCGTGGTGAATAATAATGGCAAGGCATTGTCAATTGACA 4200
QY 4200 gccaaaattggttcaaaataccatcacagactaagcaacactctagctaatgttaccaatg 4259
DB 4201 GCCAAATGTGTCAAAATACCATCACAGACTTAAGCAACACTGTAGCTAATGTTACCAATG 4260
QY 4260 ataaaggtagcgtacgcaccacagaaacagggtcaataataacaaagacagacaaaccc 4319
DB 4261 ATAAAGGTAGCGTACGCCACCACAGAACAGGCAATATTAATAAAGACAGAACAAACCC 4320
QY 4320 gtgcgcgcagcattgttgatgtgctaagcagcaggtttaacttgcgaagcgaatggtgaag 4379
DB 4321 GTGCCGCCAGCATTTGTTGATGTGCTTAAGCGCAGGCTTTAACTTGCNAGGCAATGGTGAAG 4380
QY 4380 cggttgactttgtctccacttatgacacggtcaacttggcgatggcaatgccaccacg 4439
DB 4381 CGTTTGACTTTGTCTCCACTTATGACACCGTCACTTTGCCGATGGCAATGCCACCACCG 4440
QY 4440 ctaaggtgacctatgatgacacagcaaaacagtaaaagtggtctatgatgtcaatgtgg 4499
DB 4441 CTAAGGTGACCTATGATGACACAAGCAAAACCCAGTAAAGTGGTCTATGATGTCAATGTGG 4500
QY 4500 atgatacaacattgaagttaagataaaaaacttggcgtaaaaaacccacacattgacca 4559
DB 4501 ATGATACAAACATTGAAGTTAAAGTAAATAAACTTTGGCGTAAAAACCAACCAATTGACCA 4560
QY 4560 gtactggcacaggtgtctaataattgcccataagcaatcaagctactggcgatgcgcttg 4619
DB 4561 GTACTGGCACAGGTGCTAATAATTTGCCCTTAAGCAATCAAGCTACTGGCGATGGCGCTTG 4620
QY 4620 tcaaggccagtgatcgtgtctcatctaaacaccttatctggcgacatccaaactgcca 4679
DB 4621 TCAAGGCCAGTGATCTGCTCATCTAAACACCTTTATCTGGCGACATCCAACTGCACA 4680
QY 4680 aaggggcgaagccaagcgaacaaactcagcaggtctatgtggtgctgataagggcca 4739

DB 4681 AAGGGCAAGCAAGCGAACCAACTCAGCAGGCTATGTGGATGCTGATGCCAATAAGGTCA 4740
QY 4740 tctatgacagtaccgataacaagtlactatcaagccaaaatgagtgacagttgataaaa 4799
DB 4741 TCTATGACAGTACCAGTAAACAAGTACTATCAAGCCAAAATATGATGGCACAGTTGATAAAA 4800
QY 4800 ccaagaagttgccaagacaaaactggtgcccaagcccaacccccagatggcacattgg 4859
DB 4801 CCAAGAAGTTGCCAAGACAAACTGGTCGCCCAAGCCCAAAACCCAGATGGCACATTGG 4860
QY 4860 ctcaaatgaatgtcaaatcagtcatttaacaagaacaagtaaaatgatgccataaaaaagc 4919
DB 4861 CTCAATGAATGTCAAAATCAGTCATTAAACAAGAAACAAGTAAATGATGCCAATAAANAAGC 4920
QY 4920 aaggcatcaatgaagacacaaccccttgttaaaggacttgaaaaagcccttctctgataaca 4979
DB 4921 AAGGCATCAATGAAGACAAACCGCTTTGTTAAAGGACTTGAANAAGCCCTCTCTGATAACA 4980
QY 4980 aacaaaaaacgcgcgcagtaactgtggtgatttaaatgcccgttgcacaaacacgcgtga 5039
DB 4981 AAACCAAAAACGCGCAGTAACCTGTGGGTGATTTAAATGCCGTTGCCCAACACCGCTGA 5040
QY 5040 cctttgcaagggtgatacaggcaacacgcgttaaaaaactggcgagactttgacccatcaaaag 5099
DB 5041 CCTTTGCAAGGGGTATACAGGCACAACGGCTAAAAAACTGGCGAGACTTTGACCATCAAAG 5100
QY 5100 gtggcacaacagacacccaataagctaaaccgataataaactcgggtggtgacagagtaactg 5159
DB 5101 GTGGCAAAACAGACACCAATTAAGCTTAACCGGATAATTAACATCGTGTGGTAGCAGGTACTG 5160
QY 5160 atgcttcaactgtcaaaactgtccaaagacccctaaaccaactcttaacagcgttaatgcaggtg 5219
DB 5161 ATGGCTTCACTGTCAAACTTGCCAAAAGACCTAAACCAATCTTAACAGCGTTAATGCGAGGTG 5220
QY 5220 gcacaaaattgatgacaaaaggcgtgtctttgttagactcaagcgggtcaagccaaagcaa 5279
DB 5221 GCACCAAAATTCATGACAAAGCGTGTCTTTGTAGACTCAAGCGGTCAAGCAAAAGCAA 5280
QY 5280 acacccctgtctgaagtccaatggcgtggaactgggtggcaagctcatcagtaagtgtg 5339
DB 5281 ACACCCCTGTCTAAGTCCCAATGGCGTGGACCTGGGTGGCAAGGTCAATGATGTGG 5340
QY 5340 gcaaggcacaaaagataccgcagctgccaatgtacaacagttacaacgaagtacgcaact 5399
DB 5341 GCAAAAGGCACAAAAGATACCGACGCTGCCAATGTACAAGTAAACGAAGTACGCAACT 5400
QY 5400 tgttgggtctctggtaagtgtgtaataagcgtgacgggcaatcaggttaaacattgccc 5459
DB 5401 TGTGGGTCTTTGGTAATGCTGTAATGATAAGCTTGACGGCAATCAGSTAAACATTTGCCG 5460
QY 5460 acatcaaaaaagaccccaaatcaggttcatcatctaacgcgactgtcatcaaaagcagca 5519
DB 5461 ACATCAAAAAGACCCCAATTCAGGTTTCATCTTAACCGCACTGTCTATCAAAAGCAGGCA 5520
QY 5520 cggctacttggcggtaaaggtaataacgataccgaaaaacttggcacttgggtgtatcaag 5579
DB 5521 CGGTACTTTGGCGGTAAAGGTAATAACGATACCGAAAAAACTTGCCACTGGTGTGTATACAAG 5580
QY 5580 tggcggttgataaaagacggcaacgctaacggcgatttaagcaatgttgggttcaaaaccc 5639
DB 5581 TGGCGGTGGATAAAGACGGCAACGCTTAACGGCGATTAAAGCAATGTTTGGGTCAAAACCC 5640
QY 5640 aaaaagatggcagcaaaaaagccctgtctgcacctataaagcccgaggtcagacaaact 5699
DB 5641 AAAAAGATGGCAGCAAAAAAAGCCCTGCTGCCACTTATAACGCCGCGAGGTACAGCAACT 5700
QY 5700 attgacaaacacccccagaaagccattgacagataaaatgaacaaaggtatccgcttct 5759
DB 5701 ATTTGACCAACAAACCCCGCAGAAAGCCATTGACAGAATAAATGAACAAGTATCCGCTTCT 5760
QY 5760 tccatgtcaacagatggcaatcaagagcgtgtgtacaagggcgtaacggcattgactcaa 5819

Db 5761 TCATGTCACGATGGCAATCAAGACCTGTGGTACAAAGGGCGTAACGGCATGACTCAA 5820
 Qy 5820 GTGCTCAGGCAAGCACTAGTGGGATAGTTCAGGCAAGGCAAGGATGTAAGCG 5879
 Db 5821 GTGCTCAGGCAAGCACTAGTGGGATAGTTCAGGCAAGGCAAGGATGTAAGCG 5880
 Qy 5880 CAGTGGCATAGGCAAGCAAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 5939
 Db 5881 CAGTGGCATAGGCAAGCAAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 5940
 Qy 5940 CAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 5999
 Db 5941 CAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6000
 Qy 6000 ACTGCTGCTGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6059
 Db 6001 ACTGCTGCTGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6060
 Qy 6060 ATAACAACCAAGTACCGATGCCACCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA 6119
 Db 6061 ATAACAACCAAGTACCGATGCCACCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA 6120
 Qy 6120 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6179
 Db 6121 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6180
 Qy 6180 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6239
 Db 6181 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6240
 Qy 6240 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6299
 Db 6241 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6300
 Qy 6300 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6359
 Db 6301 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6360
 Qy 6360 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6419
 Db 6361 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6420
 Qy 6420 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6479
 Db 6421 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6480
 Qy 6480 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6539
 Db 6481 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6540
 Qy 6540 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6599
 Db 6541 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6600
 Qy 6600 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6659
 Db 6601 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6660
 Qy 6660 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6719
 Db 6661 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6720
 Qy 6720 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6779
 Db 6721 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6780
 Qy 6780 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6839
 Db 6781 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6840
 Qy 6840 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6899
 Db 6841 CAGTGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 6900

Qy 6900 TGTGCGATGACCAAAAATGACCGATTTATCCGAAATTTCTGATTATGATCCGTTGA 6959
 Db 6901 TGTGCGATGACCAAAAATGACCGATTTATCCGAAATTTCTGATTATGATCCGTTGA 6960
 Qy 6960 CCGTGGATGACCAAAAATGACCGATTTATCCGAAATTTCTGATTATGATCCGTTGA 6972
 Db 6961 CCGTGGATGACCAAAAATGACCGATTTATCCGAAATTTCTGATTATGATCCGTTGA 6973
 RESULT 4
 US-08-621-944A-1
 : Sequence 1, Application US/08621944A
 : GENERAL INFORMATION:
 : APPLICANT: SASAKI, Ken
 : APPLICANT: HARKNESS, Robin E.
 : APPLICANT: LOOSMORE, Sheena M.
 : APPLICANT: KLEIN, Michel H.
 : TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
 : TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
 : NUMBER OF SEQUENCES: 10
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Sim & McBurney
 : STREET: Suite 701, 330 University Avenue
 : CITY: Toronto
 : STATE: Ontario
 : COUNTRY: Canada
 : ZIP: M5G 1R7
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/621,944A
 : FILING DATE: 26-MAR-1996
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/478,370
 : FILING DATE: 07-JUN-1995
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Stewart, Michael I
 : REGISTRATION NUMBER: 24,973
 : REFERENCE/DOCKET NUMBER: 1038-587
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (416) 595-1155
 : TELEFAX: (416) 595-1163
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 6973 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 708..6683
 : US-08-621-944A-1

Query Match 99.8%; Score 6961; DB 10; Length 6973;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 ccattgatatggcagcgtgtgctgcgcctcgcgtatgagcgagatgacacccatttgcgc 60
 Db 1 CCATGGATATGGGACAGGTGTGCTGCCCTGCGGTATGATGCGATGACACCCATTGGCC 60
 Qy 61 catatctgacgatttgacatgtgatatttaacatgtgacatgttataacattgttt 120
 Db 61 CATATCTGTACGATTTGACATGTGATGATTTAACAATGTGACATGATTTAACAATTGTTT 120

Qy	121	ataactgtgccaatcattaccataatttagtaacgacatttagtaacgcatttgtataaaat	180
Db	121	AATACTGTTGCCATCATATTACCATAATTAGTAACGCATTTAGTAACGCATTTGTAAAAAT	180
Qy	181	catgcgccccttatgtgtatcatatgaatagaatatattgatattgatctgatttgt	240
Db	181	CATTGGCCCCCTTATGTGTATCATATGAATAGAATATTAGATTGTATCTGATTATTGT	240
Qy	241	atcagaatgggtgctatataatgatgctactacgagttgatttgggttaatacactctatg	300
Db	241	ATCAGAATGGGTGATGTATATGATGATGCTTACGAGTTGATTGGGTTTAATCACTCTATG	300
Qy	301	atttgatatattttgaaactaatctatgacttaactaaatacaacatatggttataatttagca	360
Db	301	ATTTGATATATTTTGAACATAATCTATTGACTTAATCAACCATATGTTATAAATTTAGCA	360
Qy	361	taatggtagccttttggtaaaacacacatcgcaatattgtttctactgttactaccatgct	420
Db	361	TAATGTGAGCGTTTTTGTAAAAATCATATCGCAATATTGTTCTACTGTTACTACCATGCT	420
Qy	421	tgaatgacgatcccaatcaccagatttcattcaagtgtatgttttggatatacgcaccattta	480
Db	421	TGAATGAGGATCCCATCCATCCAGATTCATTCAAGTGATGCTGTTGTATACGCACCATTTA	480
Qy	481	ccctaattatttcoatacctaataatgcctatgctcagcatgtatcatcttttttaagggtaaaccac	540
Db	481	CCCTAATATTATTCAATCAAAATGCTATGTCAGCATGTATCATTTTTTTAAAGTAAACCCAC	540
Qy	541	calgaatcacatcatataaagtcattctttaacaaagccacagggcacattiatggcagtggc	600
Db	541	CATGAATTCACATCTATAAAGTATCTTTTAACAAAGCCACAGGCACATTTATGGCAGTGGC	600
Qy	601	agagtagcccaaatcccaacagcac-gggggggggtagctgtgtacagggcgaagtggcca	659
Db	601	AGAGTAGCCCAATCCACAGCACGCGGGGGGGTAGCTGTGCTACAGGCAGATTGGCA	660
Qy	560	gtgtatgacctgagccttgcggtattgcggtcgtcgtcgtcctcgtgatcgggtgcaa	719
Db	661	GTGTATGACCTGTAGCTTTGGCCGTTATTGCCGCGCTCGCTGTCTCGTGATCGGTGCAA	720
Qy	720	cgctcagtgccagtcttatgtctaaaaaaagataccaaacatatcgcaatgggtgaac	779
Db	721	CGCTCAGTGGCAGTGCTTATGCTCTAAAAAAGATACCAACATATCGCAATTGGTGAAC	780
Qy	780	aaacacagccaagcagctcaggcactgccaaaggcgagcgggtgatcgagccattgtctattg	839
Db	781	AAACACGCCAAGACGCTCAGGCACCTGCCAAGCGGCGGTGATCGAGCCATTCTCTATTG	840
Qy	840	gtgaaaaatgctaagcagcagggcggtcgaagcactcgcaatcggttagtagtaataaaactg	899
Db	841	GTGAAAATGTAAACGCACAGGCGGTTCAGGCCATCGCCATCGGTAGTAGTAATAAAAACTG	900
Qy	900	tcaatggaacagatttggataagataggtaccgatgtctacgggtcaagagtcacatcgcca	959
Db	901	TCAATGGAACAGCTTTTGGATAAGATAGTACCGATGCTTACGGGTCAAGAGTCCATCGCCA	960
Qy	960	tcggtggatgataaaggctagtggttgatgctcgtatggtccatcggttagtgatgaacttac	1019
Db	961	TCGGTGGTGATGTAAAGCGTAGTGGTGTATGCCATTCGATTCGGTAGTGATGACCTTAC	1020
Qy	1020	atttgcctgatacagcatggttaactctaaacatccgaaagggtactctctgattaacgatctta	1079
Db	1021	ATTTGCTTGATCAGCATGTTAATCTCTAAACATCCGAAAGGTACTCTGATTAACGATCTTA	1080
Qy	1080	ttaacggccaatgcagtatataaagaaaaatacgaagctcaaaagataatagtataaaaaata	1139
Db	1081	TTAACGGCCATGCAGTATTTAAAGAAATACGAAGCTCAAGAGTAATGATGTAAATATA	1140
Qy	1140	gacgcacaacccaagcgacacgcacgactactgcagtgaggcgcatgtcatatgcacagg	1199
Db	1141	GACGCACACCGCAACGGACACGCCAGTACTGTCACTGGGAGCCCATGTCTATATGCACAGG	1200
Qy	1200	gtcatatttttccaacgccttttggatcacacgggcacaacagctataaagtgctattccttggcag	1259

Db	1201	GTCAATTTTCCAAACGCTTTGGTTACACGGCCACACGCTAAAAGTGCCTATTCTCTGGCAG	1260
Qy	1260	Tgggtcttgccaccacagccgagggccaatctacaatcacttattggtctctgatgcacat	1319
Db	1261	TGGGTCTTGGCCGCCACAGCCGAGGGCCAATCTACAATTCGCTATTGGTCTCTGATGCACAT	1320
Qy	1320	ctagctcgttgggagcgtagcccttggtcaggttactcgtctcagtcacagggcagta	1379
Db	1321	CTAGCTCGTTGGAGCGATAGCCCTTGGTGCAGGTACTCGTGTCTAGTACAGGCAGTA	1380
Qy	1380	ttgccttaggtcaaggctctgttgtcaactcaagtgataataatcttagaccggcctata	1439
Db	1381	TTGGCCCTAGGTCAAGGTTCTGTGTCTACCTCAGAGTGATATAATAATCTAGACCCGGCCTATA	1440
Qy	1440	cacaaatacccgagccacttagaccccaagtctcaagccaccaataatacagaagccggctc	1499
Db	1441	CACCAAAATACCCAGGCATAGACCCCAAGTTTCAOCCACCAATAATATACGAAGCCGGGTC	1500
Qy	1500	cactttccattggttagtaactctatacaacgtaaaaatcatcattcgtgtcgaggtgtta	1559
Db	1501	CACTTTCCATTGGTAGTAACCTCTATCAACAGTAAATCATCATATGTCGGTGCAGGTGTTA	1560
Qy	1560	ataaaccgatcggttcaatgtggccacagctagaagcggtgtgtaagtggggtctaaaggac	1619
Db	1561	ATAAAACCGATCGGTCAATGTGGCACAGCTAGAAGCGGTGTGTGAAGTGGGCTTAAGGAGC	1620
Qy	1620	gtagaattacttttcagggtgatataacagctactcaactaataaaataggtttgataata	1679
Db	1621	GTAGAATTAATCTTTTCAGGGTGATGATTAACAGTACTGACGTAAAAATAGGTTTCGATTAATA	1680
Qy	1680	ctttaactattaaagtggtgcagagaccacgcattaacccgataataatcgtgtggtg	1739
Db	1681	CTTTAACTATTAAAGTGGTGCAGAGACCCACGCAITTAACCGATATAATATATCGGTGTGG	1740
Qy	1740	taaaaggagctgataatadtggtctgtaagtttaaaacttctaaactttaaaccaatotta	1799
Db	1741	TAAGAAGAGCGCTGATATAGTGTCTCGAAAGTTTAAACTTGTCTAAACCTTTAAACAACTCTTA	1800
Qy	1800	ctgaggttgatacaactcatataatgccacaccacagttaagtgtagtgtagtagta	1859
Db	1801	CTGAGGTGAATACAACCTACATTAATGCCACACACAGTTAAGGTAGGTAGTGTAGTAGTA	1860
Qy	1860	gtactacagctgaattattgagtgaagtttaacctttaccagcccaatacagggcagtc	1919
Db	1861	GTACTACAGCTGAATTTATTGAGTGATAGTTTAACCTTTACCCAGCCCAATACAGGCAGTC	1920
Qy	1920	aaagcaagcaaaaaacggtctctatggttaatgggtgaagtttactaataatgcagaaa	1979
Db	1921	AAAGCACAAAGCAAAACCGTCTATGGCGTTTAATGGGTGAAGTTTACTTAATTAATGCAGAAA	1980
Qy	1980	caacagcagaatcggcactactcgtattaccagagataaaaattggcttgcctcagatg	2039
Db	1981	CAACAGCAGCAATCGGCATCTACTCTGTTATACCAGAGATAAAAATTTGGCTTTGCTCGAGATG	2040
Qy	2040	gtgattgtgataaaaaaacgacaccataattggataaaaaaacacttaaaagtggtgtagtg	2099
Db	2041	GTGATGTTGATGAAAAACAAGCACCATATTGGATTAANAACAACCTTAAAGTGGGTAGTG	2100
Qy	2100	ttgcaattaccatagacaatggcattgtagcaggtgataaaaaagatcagtaattcttgcca	2159
Db	2101	TTTGCAATTAACCATAGACAATGGCATTTGATGCAGGTATAAAAAAGATCAGTAATCTTTGCCA	2160
Qy	2160	aaggtagcagtgtaacgatgggtttaccatgcagacgctcaaaagccgcaaaccttactt	2219
Db	2161	AAAGGTAGCAGTGTCTAACGATCGGGTTTACCATCGAACAGCTCAAAAGCCGCCAAGCCCTACTT	2220
Qy	2220	taaacgagcgctggcattcagtgacacactactgaaataatcagttgtatgctgaagtg	2279
Db	2221	TAAACGACGCGCTGGCATTGCTCACACCTTACTGAAATATCATGTTGATGCTTAAGAGTG	2280
Qy	2280	gcaatgttaccgcccacaactacaacattggcgtgaaacccacgagcttaacagtgatg	2339

Db 2281 GCAATGTTACCGCCCAACTTACAACATTGGCGTGAAAAACCAACCGAGCTTAAACAGTGATG 2340
QY 2340 gcaactagatataaatttagtgttaagggtagtgggtacgaacaataagcttagttaccgccc 2399
Db 2341 GCACCTAGTATAAATTTAGTGTAAAGGTTAGTGGTACGAACAATAGCTTAGTTACCGCGG 2400
QY 2400 aacatttggcaagctactcaaatgaagtcaatcgcaacggctgacagtgctctacaaagct 2459
Db 2401 AACATTTGGCAAGCTATCTAATAGAGTCAATCGAACGGCTGACAGTGTCTTACAAAGCT 2460
QY 2460 ttaccgttaagaagaagacgatgatgacgcccacgaogctatccacgtggcctaagaataga 2519
Db 2461 TTACCGTTAAAGAAAGACGATGATGACGCCAAACGCTATCACCGTGGCTTAAAGATACGA 2520
QY 2520 caaaaaatccgcccagctcagcatcttaaaactcaaaaggttaaaacggtctlaacggttg 2579
Db 2521 CAAAAAATGCGCGGCGAGTCAGCATCTTAAACTCAAAAGTAAACAGGTCCTAACGGTTG 2580
QY 2580 ctaccaaaaaagatggtagcgtttacaccttggccttagccaagatagcggctctgaccattg 2639
Db 2581 CTACCAAAAAAGATGTACGGTTTACCTTTGGGCTTAGCCAAAGATAGCGGCTCTGACCATTG 2640
QY 2640 gcaaaagcacccttaaacacagatggcttgactgttaagaatataccaacgaacaaatccaaag 2699
Db 2641 GCAAAAGCACCCCTAAACACGATGGCTTGACTGTAAAGATACCAACGAACAAATCCAAG 2700
QY 2700 tcggtgctaatggcattaaatttactaaatgaatggtagtaatccaggtactggcattg 2759
Db 2701 TCGGTGCTAATGCGATTAATTTACTAATGTGAATGGTAGTAATCCAGGTACTGGCATTG 2760
QY 2760 caaatccgctcgcattaccagagataaaattggcttggcttggctgtctgtatgggtgcaggtg 2819
Db 2761 CAAATACCGCTCGCATTTACCAGAGATAAAATTTGGCTTTGGTGTCTGTATGGTGGCAGTTG 2820
QY 2820 atacaaacaaaccttacttgatcaagacaagctacaagtggcaatgttaagattacca 2879
Db 2821 ATCAAAACAAACCTTATCTTGTACAAAGACAAGCTACAAGTTGGCAATGTTAAGATTACCA 2880
QY 2880 acactggcattaaacgcagggtggtaagccatcacagggtgtccccaaactgccttagca 2939
Db 2881 ACCTGCGCATTTAAGCAGGTGGTAAAGCCATCACAGGGCTGTCCCCAACACTGCCTAGCA 2940
QY 2940 ttgcgataaagttagcgcgaacatagaaactgggcaatacaatccaaagacaagacaat 2999
Db 2941 TTGCGGATCAAAAGTAGCGCAACATAGACTGGGCAATACAAATCCAAAGACAAGACAAT 3000
QY 3000 ccaacgctgcagcatttaattgatataataatcacaggtttaaacttaaaataaataaaca 3059
Db 3001 CCAACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAACCCTAAAAAATAATAACA 3060
QY 3060 accccattgacttgtctccacttatgacattgttgacttggcaatggcaatggcaacga 3119
Db 3061 ACCCATTTGACTTTGTCTCCACTTATGACATTTGTGACTTTGGCAATGGCAATGGCCACCA 3120
QY 3120 ccgcacagtaaccatgataccgtcaacaaaccagtaagtggtatatgtatgtggaatg 3179
Db 3121 CGGCACNGTRAACCCATGATACCCGCTAACAAACCAAGTAAAGTGGTATATGATGTGAATG 3180
QY 3180 tggatgatataccatttcaatacaggcactgatgacaaataaaaaaacttggcgctcaaaa 3239
Db 3181 TGGATGATACAACCATTCATTAACAGGCACTGATGACATTAATAAATAAATAAATAAATAA 3240
QY 3240 ccacaaaactgaacaaaacgaatgctaatgtaatacagcaactaaacttlaattgtaact 3299
Db 3241 CCACCAAACTGAACAAACAAGTGTAAATGGTAATACAGCACTTAACTTTAATGTTAACT 3300
QY 3300 ctagtgaagaatgccccttggtaacgccaagacatcgccgaataatcctaaccacctag 3359
Db 3301 CTAGTGATGAAGATGCCCTTGTAAAGCCAAAGACATCGCCGAAAAATCTTAAACACCCTAG 3360
QY 3360 ccaaggaaattcacaccccaaaaggcacagacacccgcccctcaaaacctttaccgtta 3419
Db 3361 CCAAGGAAATTCACACCACCAAGGACAGACACCGCGCCTACAAACCTTTTACCCTTA 3420

QY 3420 aaaaggtagatgaataataatgtctgatgacgcgaacgccatcaccggtgggtcaaaaga 3479
Db 3421 AAAAGGTAGATGAATAATAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3480
QY 3480 acgcaataatacaagtcacacccctcaacactcaaaagtgcaaaacggcttcaataattaaaa 3539
Db 3481 ACGCAAAATTAATCAAGTCAACACCCCTAACACTCAAAAGGTGAAACGGTCTTAAATATTAAA 3540
QY 3540 cgcacaaaatgggtacggttacgtttacgttggcattacacacacgaagcggcttaaaagccggca 3599
Db 3541 CGACAAAAATGCTACGCTTACCTTTGGCATTAACACCAACAGCGGTCTTAAAGCCGGCA 3600
QY 3600 aaagcaccctaaagacgggtggtctgtctataaaaaacccacactggtagcgaacaaatcc 3659
Db 3601 AAAGCACCTTAACAGCGGTGGCTTGTCTATTAAAAACCCCACTGGTAGCAGCAANAATCC 3660
QY 3660 aagtcgggtgctgatggcgtgaagtttgcgaagtttaataataatgggttcttagtgctg 3719
Db 3661 AAGTCGGTGTGATGGCGTGAAGTTTGGCAAGGTTAAATAAATAGTGTGTTAGTGGTGTG 3720
QY 3720 gcattgatggcacaaactcgcattaccagagatgaatttggctttactgggactaatggct 3779
Db 3721 GCATTTGATGGCACAACTCGCATTAACAGAGATGAATAATTTGGCTTTTACTGGGACTAATGGCT 3780
QY 3780 cacttgaataaaagcaaaacccacactaaagcaagacggcatttaacgcaggtggttaaaaga 3839
Db 3781 CACTTGTATAAAGCAAAACCCCACTTAAGCAAAAGACGGCATTAACGACAGGTGGTAAAAAGA 3840
QY 3840 ttaccaacattcaatcaggtgagattggcccaaaacagcagcagtgctgtgacagcgccga 3899
Db 3841 TTACCAACAATTCATCAGGTGAGATTGCCCAAAACAGCCATGATGCTGTGACAGCGGCA 3900
QY 3900 agatttatgatttaaaacccgaacttgaacaaacaaatcagcagtagctgccaaaaacagcac 3959
Db 3901 AGATTTATGATTTAAAAACCGAACTTGAACAAATAACAGCACTACTGCCAAAACAGCAC 3960
QY 3960 aaactcattacacaattctcagtagcagatgaacaggttaacttaacttcaggttagca 4019
Db 3961 AAAACTCATTAACAGAAATTTCTCAGTAGCAGATGAACAGGTAACTTAACCTTACGTTAGTA 4020
QY 4020 accttactccagtttatgacacctcaagacccctcgtgatgtcatcacttgcaggtgaaa 4079
Db 4021 ACCCTTACTCCAGTTATGACACTCAAGACCTCTGATGTCATCACTTTTGCAGGTGAAA 4080
QY 4080 acggcattaccaccaaggttaataaaaggtgtggtgctgtggcatttgaccaaaacaaag 4139
Db 4081 ACGGCATTACCAACCAAGGTAATAAAGGTGTGGTGGGTGGGCTATGACCAAAACCAAG 4140
QY 4140 gcttaaccgccttaagctgacggtgggttaataataatggcaaaagcattgtcattgaca 4199
Db 4141 GCTTAACCAAGCTAAGCTGACCGTGGTAAATAATAATGGCAAGGCAATGTCATTTGACA 4200
QY 4200 gccaaaatgggtcaaaaataccatacagggactaagcaaacctctagtaattgttaccacag 4259
Db 4201 GCCAAAATGGTCAAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTTACCAGTG 4260
QY 4260 ataaaggtagctgacacacacagacagggcaataataataaaagcgaagacaaaaccc 4319
Db 4261 ATAAAGGTAGCTACGCCACACAGACAGGGCAATATATCAAAAGACGAAGACAAAACCC 4320
QY 4320 gtgcccacagcattgttgatgtgctgaagcagcgtttaaacttgaaggaatggggaag 4379
Db 4321 GTCCCGCCAGCATTTGTGATGTGCTAAGCGCAGGCTTAACTTGCAGGCAATGTTGGTGAAG 4380
QY 4380 cggttgactttgtctccacttatgacacgtcaacttccgatgcaatgcccaccacg 4439
Db 4381 CGGTTGACTTTGTCTCCCTTATGACACCGCTCAACTTTGCCGATGCAATGCCACCAACCG 4440
QY 4440 ctaaggtgacattatgacacacgaacaaacccagtaaaagtggtctatgatgtcaatgtgg 4499
Db 4441 CTAAGGTGACCTATGATGACACAAAGCAAAACCAAGTAAAGTGGTCTATGATGTCATGTGG 4500

Qy	1080	ttaacggccatgcagttattaaagaaatacagaagctcaagagataatgattgtaaaatata	1139
Db	1081	TTAACGGCCATGCAGTATTAAAAGAAATACGAAGCTCAAGAGTAATGATGATAAATATA	1140
Qy	1140	gacgcacaaccgaagcgagacacccagttactgcagtgaggagccatctcatatgcacagg	1199
Db	1141	GACGCACACCCGACAGCGGACACGCCAGTACTGCAGTGGGAGCCATGTCATATGCACAGG	1200
Qy	1200	gtcatatttccaacgcctttggctacacggggcaacagctaaagtgccatttcccttggcag	1259
Db	1201	GTCATTTTCCAAAGCCTTTGGTACACGGCAACAGCTAAAAGTGCTATTCTCTGGCAG	1260
Qy	1260	tgggtcttgccgccacagccgagggcgaatctacaaatcgctatttggttcttgatgcacaat	1319
Db	1261	TGGGTCTTGCCGCCACAGCCGAGGGCCAAATCTACAATCGCTATTGGTTCTTGATGCAACAT	1320
Qy	1320	ctagctcgtttgggagcgatagcccttggctgcaggtactcgtctgcagctcacaggcgagta	1379
Db	1321	CTAGCTCGTTGGAGCGATAGCCCTTGGTGCAGGTACTCTGTCGTCTACGCTACAGGGCAGTA	1380
Qy	1380	ttgccctaggtcaaagttcgtgttcactcagagtgataataattctagaccggccctata	1439
Db	1381	TTGCCCTAGGTCAAGGTTCTGTGTCTACTCAGAGTGATAATAATTCCTAGACCGGCTATA	1440
Qy	1440	caccaataaccagcactagaccccccaagtttcaagccaccaataatacagaagcggggtc	1499
Db	1441	CACCAATAACCCAGGCACTAGACCCCAAGTTTCAAGCCACCAATATATACGAAGCGGGTC	1500
Qy	1500	cactttccattggtagtaactctatacaacggtataaatacatcaatgctcgttgccaggtgtta	1559
Db	1501	CACTTTCATTTGGTAGTAACTCTATCAACGTTAAATCATCAATGTCGGTGCAGGTGTTA	1560
Qy	1560	ataaaaccgatcggttcaaagtggccacagctagaagcgggtggtaagtgggctaaaggagc	1619
Db	1561	ATAAAACCGATCGGTCAATGTGGCACAGCTAGAAAGCGGTGGTCAAGTGGGCTAAGGAGC	1620
Qy	1620	gtagaattactttcaagtgatgaataacagtactgcagtaaaataaggtttgtgataata	1679
Db	1621	GTAGAATTACTTTTCAGGTGATGATACAGTACTGCGTAAAAATAGGTTTGGATAATA	1680
Qy	1680	ctttaactataaagtggtgcagagaccaacgcattaaccgataataatactcgggtgtgg	1739
Db	1681	CTTTAACTATTAAAGTGTGTCAGAGACCAACGCATTAAACCGATAATAATATCGGTCTGG	1740
Qy	1740	taaaagagcctgataatadggcttgaagttaaacttgctaaactttaacaacttta	1799
Db	1741	TAAAGAGGCTGATAATAGTGGTCTGAAAGTTAAACTTTGCTTAAACCTTTAAACAACTTTA	1800
Qy	1800	ctgagtgtaatacaactacattaaatgcccacaacacagtttaaggttaggtagtagtagta	1859
Db	1801	CTGAGGTGAATACACTACATTAAATGCCACACCACAGTTTAAGGTTAGGTAGTAGTAGTA	1860
Qy	1860	gtactacagctgaattattgagtgtagtttaaacctttaccgcccccaatacagggcagtc	1919
Db	1861	GTACTACAGCTGAATTATTGAGTGATGTTTAACTTTTACCCGCCCAATACAGGCAGTC	1920
Qy	1920	aagcacaagcaaaacgcctctatggcgttaatbggggtgaagtttacttaataatgcagaaa	1979
Db	1921	AAAGCAACAAGCAAAACCGTCTATGGCGTTATGGGGTGAAAGTTTACTTAATATCCAGAAA	1980
Qy	1980	caacagcgaactcggcactactcgtattaccagagataaaatggcctttgcttcgagatg	2039
Db	1981	CACAGCAGCAATCGGCATCTACTCGTATTACAGAGATAAANTTGGCTTTGCTCGAGATG	2040
Qy	2040	gtgatgttgatgaaaaaacgaccataatttggataaaaaaacaacttaaaagttgggtagtg	2099
Db	2041	GTGATGTTGATGAAAAACAAGCACCATATTTGGATAAAAAACAACCTTTAAAGTGGGTAGTG	2100
Qy	2100	ttgcaattaccatagacaaatggcatatgatgcaggttaataaaaaagatcagtaattctgcca	2159
Db	2101	TTGCAATTACCATAGACAATGGCATTTGATGCAGGTATATAAAGATCAGTAATCTTTGCCA	2160

QY	2160	aaggtagcagtgctaacgatgcggtttaccatcgacagctcaagcgcgcaagcctactctt	2219
DB	2161	AAAGGTAGCAGTGCTAACGATGGGTTTACCATCGAACAGCTCAAAAGCGCGCAAGCCTACTT	2220
QY	2220	taaacqcgaggcgctggcatcagtgctcacacctactgaaatacatcagttgatgctaaagtg	2279
DB	2221	TAAACCGAGCGCTGGCATCAGTGTCACACCTACTGAAATATACAGTGTAGTGCCTAAAGAGTG	2380
QY	2280	gcaatgttaccgccccaaacttacaacatggcggtgaaacaccacggagcttaacagtgatg	2339
DB	2281	GCAATGTTACCGCCCACTTACACATTGGCGTGAACACCACCGAGCTTAAACAGTGATG	2340
QY	2340	gcactagtataaattagtgtaagggttagtgggtacgaacaatagcttagttaccgcgg	2399
DB	2341	GCACTAGTGATAAATTTAGTGTTAAGGGTAGTGGTACGAACAATAGCTTAGTTACCGCGC	2400
QY	2400	aacatttggcgaagctatcaaatgaagtcaatcgaaaggctgacaagtgccttcaaaagct	2459
DB	2401	AACATTTGGCAAGCTATCTAAATGAAGTCAATCGAAGCGCTGACAGTGCTCTACAAGCT	2460
QY	2460	ttaccggttaaaagaagaagacgatgatgcgcgaacgctatcacctgggcttaaaagatacga	2519
DB	2461	TTACCGTTAACGAGNAGACGATGATGAGCGCCAACGCTATCACCGTGGCTTAAGATACGA	2520
QY	2520	caaaaatgcgcgcgcagtcagcatcttaaaactcaaggttaaaaacgggtcttaacgggtg	2579
DB	2521	CAAAAAATGCGCGCAGTCAGCATCTTAAAACTCAAAGTAAAAACGGGTCTTAACGGTTG	2580
QY	2580	ctaccaaaaaagatggtacgggttacctcttgggcttagcaagatagcgtctgaaccattg	2639
DB	2581	CTACCAAAAAAGATGGTACGGTTACCTTTTGGCGTTTAGCCAAGATACGGGTCTGCACCATG	2640
QY	2640	gcaaaagcacccctaaacaacgatggcttgactgttaaagataccaacgaacaaatccaag	2699
DB	2641	GCAAAAGCACCCCTAACACACGATGGCTTGACTGTTAAAGNATCCACAGCACAANTCCAAG	2700
QY	2700	tcgggtgtaatggcattaaattactaatgtgaatggttagtataccaggtaactgggcattg	2759
DB	2701	TCGGTGCTAATGGCATTAATTTACTTAATGTGAATGGTAGTAATCCAGGTACTTGGCATTG	2760
QY	2760	caaatcacgcctcgacttaccagagataaaatggcttcttgctgtctgatggtgcagttg	2819
DB	2761	CAAAATACCGCTCGCATACAGAGATAAAATTTGGCTTTGCTGGTCTGTAGTGGTGACAGTTG	2820
QY	2820	atacaaaaacacttatcttgatacaagaacgatacaagtgtggcaatgttaaagattacca	2879
DB	2821	ATACAAACAACACCTTATCTTGATCAGACACAGCTACAGTTGGCATGTTAAGATTACCA	2880
QY	2880	acactggccattaacgcagtggttaagccatcacagggtgtccccaaactgccttagca	2939
DB	2881	ACACTGGCATTAACGCAGGTGGTAAAGCCATCACAGGGGTGTCCCCAACACTGCCTAGCA	2940
QY	2940	ttgcgcgatacaagtagccgcaacatagaaactgggcgaataacaatccaagacaaagacaaat	2999
DB	2941	TTTGGCCGATCAAAAGTAGCCCAACATAGAACTGGGCCAATAACAATCCAAGACAAAGACAAAT	3000
QY	3000	ccaacgctccagcattaatgatataataaacagcctttaacctaaaaataaataaaca	3059
DB	3001	CCACAGCTGCCAGCATTAATGATATATTAATAACAGCGCTTAACCTTAAAAAATAATNACA	3060
QY	3060	acccattgactttgctccacttatgacattgttgactttgccaattgccaatggcaatgccacca	3119
DB	3061	ACCCCATTGACTTTGTCTCCACTTATGACATTGTTGACTTTTGGCAATGGCAATGCCACCA	3120
QY	3120	cgcgcacagtaacccaatgataccgcttaacaaaaccagtaaaagtgggtatatgatgtgaatg	3179
DB	3121	CGCGCACAGTAACCCATGATACCGCTTAACAAAAACCCAGTAAAGTGTATATGATGTGAATG	3180
QY	3180	tgatgatatacaaccattcatctaaacaggcactgatgacataaaaaaacttggcgctcaaaa	3239
DB	3181	TGGATGATACACCAATTCATCTAACAGGCACTGATGACAAATAAAAAAATTTGGCGCTCAAAA	3240
QY	3240	ccacaaactgacaaaaacaagtgcgtaattggttaatacacgacaaacttaattgttaact	3299

Qy 1 ccattggatatgggcagggtgtgctcgcctgcgctatgatggcgatgacacccatttgc 60

Db 1 ccatggatgaggcaggtgtgtctgcgcgtatgtatggcgatgacaccccatgtggcc 60
Qy 61 catatctgacgattgacatgtgatattgaatcatgtgacatgatttaacattgttt 120
Db 61 catatctgacgattgacatgtgatattgaatcatgtgacatgatttaacattgttt 120
Qy 121 aatctgttgccatcattaccataatttagtaaacgatttagtaacgcatattgtaaaat 180
Db 121 aatctgttgccatcattaccataatttagtaaacgatttagtaacgcatattgtaaaat 180
Qy 181 catgtgcgcctttatgtgtatcatatagaatagaatattatgattgattgatttattgt 240
Db 181 catgtgcgcctttatgtgtatcatatagaatagaatattatgattgattgatttattgt 240
Qy 241 atcagaatggtagctatattgtagtgcctacagattgattgggttaattcaactctatg 300
Db 241 atcagaatggtagctatattgtagtgcctacagattgattgggttaattcaactctatg 300
Qy 301 atttgatattttgaaactaatctattgactttaataaccacatattggtataatttagca 360
Db 301 atttgatattttgaaactaatctattgactttaataaccacatattggtataatttagca 360
Qy 361 taatgtgagctttttgtaaaatcacatgcgaatttcttactgtttactacacatgct 420
Db 361 taatgtgagctttttgtaaaatcacatgcgaatttcttactgtttactacacatgct 420
Qy 421 tgaatgacgattcccaatcaccagatttcaagtgatgtgtttgtatcacgacatttta 480
Db 421 tgaatgacgattcccaatcaccagatttcaagtgatgtgtttgtatcacgacatttta 480
Qy 481 cccataattatttcaaatgcctatgtatgcagtgatcatctttttaaaggtaaacacac 540
Db 481 cccataattatttcaaatgcctatgtatgcagtgatcatctttttaaaggtaaacacac 540
Qy 541 catgaatcacatctataaagtcatttcaacaaagccacagcacatttatggcagtgcc 600
Db 541 catgaatcacatctataaagtcatttcaacaaagccacagcacatttatggcagtgcc 600
Qy 601 agatgacgaccccaatccacagcac-gggggggggtagctgtgctacagggcaattgca 659
Db 601 agatgacgaccccaatccacagcacggggggggtagctgtgctacagggcaattgca 660
Qy 660 gtgtatgacactgagctttgcccgatttgcgcgctgcgtctcctgtatgctgtgcaa 719
Db 661 gtgtatgacactgagctttgcccgatttgcgcgctgcgtctcctgtatgctgtgcaa 720
Qy 720 cgtcagtgagctgtcttatgtctcaaaaaaagataaccaaacatatacgaattggaac 779
Db 721 cgtcagtgagctgtcttatgtctcaaaaaaagataaccaaacatatacgaattggaac 780
Qy 780 aaaaacagcgaagcgtcaggaactgccaaggcagcggtagctgagccattgctattg 839
Db 781 aaaaacagcgaagcgtcaggaactgccaaggcagcggtagctgagccattgctattg 840
Qy 840 gtgaataatgctaacacagggcgtcaagccatcgccatcggttagtgaataaaactg 899
Db 841 gtgaataatgctaacacagggcgtcaagccatcgccatcggttagtgaataaaactg 900
Qy 900 tcaatggaagcagtttggataagataggtaccgatgctacgggtcaagagtcacatcgcca 959
Db 901 tcaatggaagcagtttggataagataggtaccgatgctacgggtcaagagtcacatcgcca 960
Qy 960 tgggtggtagtgaagcgtcagtgatgctcgcagatgccaatcggttagtgacttac 1019
Db 961 tgggtggtagtgaagcgtcagtgatgctcgcagatgccaatcggttagtgacttac 1020
Qy 1020 atttgcctgacgacatggttaactctaaacatccgaaggctactctgattacgacttta 1079
Db 1021 atttgcctgacgacatggttaactctaaacatccgaagggtactctgattacgacttta 1080
Qy 1080 ttaacggccatgagctatttaaaagaaatcacgaagctcaaaagataatgatgtaaataata 1139
Db 1081 ttaacggccatgagctatttaaaagaaatcacgaagctcaaaagataatgatgtaaataata 1140

Qy 1140 gacgcacaacgcgaagcgcacacgcagctactgcagttgggagccatgtcataatgcacagg 1199
Db 1141 gacgcacaacgcgaagcgcacacgcagctactgcagttgggagccatgtcataatgcacagg 1200
Qy 1200 gtcatttttccaaacgcctttgtgtacacggaacagctaaagtccttacttcttggcag 1259
Db 1201 gtcatttttccaaacgcctttgtgtacacggaacagctaaagtccttacttcttggcag 1260
Qy 1260 tgggtcttgcgcacacagcgcagggccaatctacaatcgctatttgggtctgagcaacat 1319
Db 1261 tgggtcttgcgcacacagcgcagggccaatctacaatcgctatttgggtctgagcaacat 1320
Qy 1320 ctactgcgttgggagcgtatgccttgggtgacaggttactctgctcagctacagggcagta 1379
Db 1321 ctactgcgttgggagcgtatgccttgggtgacaggttactctgctcagctacagggcagta 1380
Qy 1380 ttgcctagggtcaaggttctgttgcactcagagtgataataattctagacggcctata 1439
Db 1381 ttgcctagggtcaaggttctgttgcactcagagtgataataattctagacggcctata 1440
Qy 1440 caccaaaatcaccagcactagaccccaagttccaagccaccaataatcgaagggcggctc 1499
Db 1441 caccaaaatcaccagcactagaccccaagttccaagccaccaataatcgaagggcggctc 1500
Qy 1500 cactttccatttggtagtaactctatacacaagtcataaataatcactgcgtgagtggtta 1559
Db 1501 cactttccatttggtagtaactctatacacaagtcataaataatcactgcgtgagtggtta 1560
Qy 1560 ataaaacgcagtcggttcaatgtgacacagctagaagcgtgtggaagtgggctaaaggagc 1619
Db 1561 ataaaacgcagtcggttcaatgtgacacagctagaagcgtgtggaagtgggctaaaggagc 1620
Qy 1620 gtagaatttacttttcaggggtgatataacagctactgacgttaaaaatagatttggataata 1679
Db 1621 gtagaatttacttttcaggggtgatataacagctactgacgttaaaaatagatttggataata 1680
Qy 1680 ctttaactattaaaagtggtgacagagcaacgcatttaacocgataataataatcgggtgagg 1739
Db 1681 ctttaactattaaaagtggtgacagagcaacgcatttaacocgataataataatcgggtgagg 1740
Qy 1740 taaaagagcgtgataatagtggtctgaaagttaaacttgcataaaactttaaaacacttta 1799
Db 1741 taaaagagcgtgataatagtggtctgaaagttaaacttgcataaaactttaaaacacttta 1800
Qy 1800 ctgaggtgaatacaactacattaaatgcacaaacacacagtttaaggtaggtagtagta 1859
Db 1801 ctgaggtgaatacaactacattaaatgcacaaacacacagtttaaggtaggtagtagta 1860
Qy 1860 gtactacagctgaattatttagtgatgattttaacactttaccagcccaatacagggcagtc 1919
Db 1861 gtactacagctgaattatttagtgatgattttaacactttaccagcccaatacagggcagtc 1920
Qy 1920 aagcacaagaacaaacccgtctatggcgttaattgggggtgaaagttaactaataatgcagaaa 1979
Db 1921 aagcacaagaacaaacccgtctatggcgttaattgggggtgaaagttaactaataatgcagaaa 1980
Qy 1980 caacagcgaactcggcactactctattaccagagataaaaattgggttgcctgcagagtcg 2039
Db 1981 caacagcgaactcggcactactctattaccagagataaaaattgggttgcctgcagagtcg 2040
Qy 2040 gtgattgtgataaaacaaacagcaccatatttggataaaaaaacacttaaaagtgggtagtcg 2099
Db 2041 gtgattgtgataaaacaaacagcaccatatttggataaaaaaacacttaaaagtgggtagtcg 2100
Qy 2100 ttgcaattaccatagacaatggcattgtgcaggttaataaaaaagatcgactaactcttgcca 2159
Db 2101 ttgcaattaccatagacaatggcattgtgcaggttaataaaaaagatcgactaactcttgcca 2160
Qy 2160 aaggtagcagtgctaacgatgcggttaccatcgacagctcaaaagccgcaagcctactt 2219
Db 2161 aaggtagcagtgctaacgatgcggttaccatcgacagctcaaaagccgcaagcctactt 2220

Db 4381 cgggtgacttggctccacttatgacacggtcaactttggcgaatggcaatgccaccacg 4440
Qy 4440 ctaaggtgacctatgatgacacaaagcaaaacagtaaaagtgggtcttatgatgtcaatgtg 4499
Db 4441 ctaaggtgacctatgatgacacaaagcaaaacagtaaaagtgggtcttatgatgtcaatgtg 4500
Qy 4500 atgatacaacattgaagttaaaagataaaaaacttggcgttaaaacaccaccatttgacca 4559
Db 4501 atgatacaacattgaagttaaaagataaaaaacttggcgttaaaacaccaccatttgacca 4560
Qy 4560 gtactggcacaggtgcttaataatttggcctaagcaatcaagctactggtgcatggtg 4619
Db 4561 gtactggcacaggtgcttaataatttggcctaagcaatcaagctactggtgcatggtg 4620
Qy 4620 tcaagggcagtgatgtctgttgcctactataacacotttatctggtgcacatccaaactgcca 4679
Db 4621 tcaagggcagtgatgtctgttgcctactataacacotttatctggtgcacatccaaactgcca 4680
Qy 4680 aaggggcaagccaagcaaacactcagcaggtctatgtggtgctgtggtgcaataaaggtca 4739
Db 4681 aaggggcaagccaagcaaacactcagcaggtctatgtggtgctgtggtgcaataaaggtca 4740
Qy 4740 tctatgacagtgatgataacaaagtaactatcaagggccaataatgatggcagttgtataaa 4799
Db 4741 tctatgacagtgatgataacaaagtaactatcaagggccaataatgatggcagttgtataaa 4800
Qy 4800 ccaaaagaattgccaagaagcaaacgtgtgcccagcccaagcccaaacccagatggcacattgg 4859
Db 4801 ccaaaagaattgccaagaagcaaacgtgtgcccagcccaagcccaaacccagatggcacattgg 4860
Qy 4860 ctcaaatgaatgctcaaatcagtcatttaacaaagaaacaaagtaaatgatggccaataaaagc 4919
Db 4861 ctcaaatgaatgctcaaatcagtcatttaacaaagaaacaaagtaaatgatggccaataaaagc 4920
Qy 4920 aaggtcatatgaagcaacgcgttgggttaagaggacttgaagaagccgttctgtataaca 4979
Db 4921 aaggtcatatgaagcaacgcgttgggttaagaggacttgaagaagccgttctgtataaca 4980
Qy 4980 aaacccaaacgcgcagtaactgtgggtgatttaaatccgttggcccaaacacccgtga 5039
Db 4981 aaacccaaacgcgcagtaactgtgggtgatttaaatccgttggcccaaacacccgtga 5040
Qy 5040 ccttgcaggggatacaggcacaacgcgttgaataaaactggggcagagacttggaccataaag 5099
Db 5041 ccttgcaggggatacaggcacaacgcgttgaataaaactggggcagagacttggaccataaag 5100
Qy 5100 gtgggcaaacagacacaaataagctaacgcgataataacatcggtgtgggtagcaggtactg 5159
Db 5101 gtgggcaaacagacacaaataagctaacgcgataataacatcggtgtgggtagcaggtactg 5160
Qy 5160 atggcttcaactgtcaaaacttgcacaaagacctaaccatacttaacagcgttaatgcagggtg 5219
Db 5161 atggcttcaactgtcaaaacttgcacaaagacctaaccatacttaacagcgttaatgcagggtg 5220
Qy 5220 gcaccaaattgatgacaaaggcgtgtcttctttagactcaagcgttcaagccaaagcaa 5279
Db 5221 gcaccaaattgatgacaaaggcgtgtcttctttagactcaagcgttcaagccaaagcaa 5280
Qy 5280 acacccctgtctaaagtgcacaaatgggtgacctggtggcgaagggtcaatcagtaattgtgg 5339
Db 5281 acacccctgtctaaagtgcacaaatgggtgacctggtggcgaagggtcaatcagtaattgtgg 5340
Qy 5340 gcaaaagcacaacaaagataccgcagcgtccaatgtacacaggttaaacgaagtacgcaact 5399
Db 5341 gcaaaagcacaacaaagataccgcagcgtccaatgtacacaggttaaacgaagtacgcaact 5400
Qy 5400 tgttgggtcttggtaagtgtctggtaagtataacaggttaaacgaagtacgcaact 5459
Db 5401 tgttgggtcttggtaagtgtctggtaagtataacaggttaaacgaagtacgcaact 5460
Qy 5460 acatcaaaaaagaccccaattcaggttcaatcaatcaacccgactgtctcaaaagcagga 5519
Db 5461 acatcaaaaaagaccccaattcaggttcaatcaatcaacccgactgtctcaaaagcagga 5520

Qy 5520 cgggtacttggcgttaaaaggttaataacgataccgaaaaacttgcactgggtgggtatacaag 5579
Db 5521 cgggtacttggcgttaaaaggttaataacgataccgaaaaacttgcactgggtgggtatacaag 5580
Qy 5580 tgggcgttggataaaagacggaacgctaacgagcgttttaagcaatgttttgggtcaaaaaccc 5639
Db 5581 tgggcgttggataaaagacggaacgctaacgagcgttttaagcaatgttttgggtcaaaaaccc 5640
Qy 5640 aaaaagatggcagcaaaaaaacgctgtccacttataacgcccgcaggtcagaccacact 5699
Db 5641 aaaaagatggcagcaaaaaaacgctgtccacttataacgcccgcaggtcagaccacact 5700
Qy 5700 atttgaccaacaaacccgcagagccatttgacagataaaatgaacaaaggttatccgcttct 5759
Db 5701 atttgaccaacaaacccgcagagccatttgacagataaaatgaacaaaggttatccgcttct 5760
Qy 5760 tccatgtcaacgataggcaaatcaagagcctgttgggtacaaggcgttaacggcgaattacacaa 5819
Db 5761 tccatgtcaacgataggcaaatcaagagcctgttgggtacaaggcgttaacggcgaattacacaa 5820
Qy 5820 gtgcctcaggaacgacactcagtgccgatagttttccaggccaagcagatggtgaagccg 5879
Db 5821 gtgcctcaggaacgacactcagtgccgatagttttccaggccaagcagatggtgaagccg 5880
Qy 5880 cgttggcattaggcagacaaaccccaagcaggcaacccaattccatcgccatcggtgataaagc 5939
Db 5881 cgttggcattaggcagacaaaccccaagcaggcaacccaattccatcgccatcggtgataaagc 5940
Qy 5940 cacaagcacggcgatcaatccatcgccatcggttacaggccaatgttggtagcaggtaaagc 5999
Db 5941 cacaagcacggcgatcaatccatcgccatcggttacaggccaatgttggtagcaggtaaagc 6000
Qy 6000 actctggtgccatcgccagcccaacgactgtttaagcctgataacagttcacagttcgggta 6059
Db 6001 actctggtgccatcgccagcccaacgactgtttaagcctgataacagttcacagttcgggta 6060
Qy 6060 ataaacacagtttaccgatgccactcaaacgatgttttgggtggtggcaataacatca 6119
Db 6061 ataaacacagtttaccgatgccactcaaacgatgttttgggtggtggcaataacatca 6120
Qy 6120 ccgtgaccgaaagttaactcgttgccttaggttcaaaactctgccatcgatgtaggacacac 6179
Db 6121 ccgtgaccgaaagttaactcgttgccttaggttcaaaactctgccatcgatgtaggacacac 6180
Qy 6180 acgaggcacacaagcccaaaaaatctgcagcgacacagcaggtacacaccacacagcaggtg 6239
Db 6181 acgaggcacacaagcccaaaaaatctgcagcgacacagcaggtacacaccacacagcaggtg 6240
Qy 6240 caacgggtacggttaaaagcgttgcctgggacaaacgcgggttgggtcctcgttgggtg 6299
Db 6241 caacgggtacggttaaaagcgttgcctgggacaaacgcgggttgggtcctcgttgggtg 6300
Qy 6300 cctcaggtgtgaaacgcgttataccaaaaatgtgggacagcaggttaggttcagtcaccacagca 6359
Db 6301 cctcaggtgtgaaacgcgttataccaaaaatgtgggacagcaggttaggttcagtcaccacagca 6360
Qy 6360 cagatcggtcaatgttagccaggtttatacaagcccaacaaagcattgccaaacgcaacca 6419
Db 6361 cagatcggtcaatgttagccaggtttatacaagcccaacaaagcattgccaaacgcaacca 6420
Qy 6420 atgagcttgacctgtatccacaaaacgaaaaataaggccaatgagggatttctcag 6479
Db 6421 atgagcttgacctgtatccacaaaacgaaaaataaggccaatgagggatttctcag 6480
Qy 6480 cagtgccgatggcgctccatggcacaagcctatacttctggcagatccatgggtaccgggg 6539
Db 6481 cagtgccgatggcgctccatggcacaagcctatacttctggcagatccatgggtaccgggg 6540
Qy 6540 gtattgccaccacacacggttcaaggttgcggtggcagttgggactgtcgaagcttcgggata 6599
Db 6541 gtattgccaccacacacacggttcaaggttgcggtggcagttgggactgtcgaagcttcgggata 6600

Qy	1200	gtcattttccaaagcccttgggtacacgggcaacagctaaagtgctattccttgccag	1259
Db	1201	gtcattttccaaagcccttgggtacacgggcaacagctaaagtgctattccttgccag	1260
Qy	1260	tgggtcttgccgccacagccgagggccaatcacatcgctattggttctgtatgcacat	1319
Db	1261	tgggtcttgccgccacagccgagggccaatcacatcgctattggttctgtatgcacat	1320
Qy	1320	ctagctcgttgggagcgatagcccttgggtgaggtactcgtgctcagctacagggcagta	1379
Db	1321	ctagctcgttgggagcgatagcccttgggtgaggtactcgtgctcagctacagggcagta	1380
Qy	1380	ttgccctaggtcaaagttctgtgtccactcagagtgcataatactctagaccggcctata	1439
Db	1381	ttgccctaggtcaaagttctgtgtccactcagagtgcataatactctagaccggcctata	1440
Qy	1440	caccaataccagcgactagaccccaagtttcaagccaccaataaacacgaagcgggtc	1499
Db	1441	caccaataccagcgactagaccccaagtttcaagccaccaataaacacgaagcgggtc	1500
Qy	1500	cactttccattggtagtaactctataaacgttaaaatacctaagtcggtgcaggtgtta	1559
Db	1501	cactttccattggtagtaactctataaacgttaaaatacctaagtcggtgcaggtgtta	1560
Qy	1560	ataaaaccgatcggtccaatgtggcacagctagaaagcgggtggtgaagtgggctlaaagagc	1619
Db	1561	ataaaaccgatcggtccaatgtggcacagctagaaagcgggtggtgaagtgggctlaaagagc	1620
Qy	1620	gtagaattacttttcagggtgatataacagttactgcgtataaaataagggtttggataata	1679
Db	1621	gtagaattacttttcagggtgatataacagttactgcgtataaaataagggtttggataata	1680
Qy	1680	ctttaactattaagtggtgcagagaccaacgcgattaaccgataataatactcgggtgtgg	1739
Db	1681	ctttaactattaagtggtgcagagaccaacgcgattaaccgataataatactcgggtgtgg	1740
Qy	1740	taaaagggctgataatagtggtctgaaagtttaaaacttgcataaacttttaaaacaatttta	1799
Db	1741	taaaagggctgataatagtggtctgaaagtttaaaacttgcataaacttttaaaacaatttta	1800
Qy	1800	ctgagggtgaatacaactacataataatgacacacacagtttaaggttaggtagtagta	1859
Db	1801	ctgagggtgaatacaactacataataatgacacacacagtttaaggttaggtagtagta	1860
Qy	1860	gtactacagctgaattattgagttagtttaacctttaccagccccaatacagggcagtc	1919
Db	1861	gtactacagctgaattattgagttagtttaacctttaccagccccaatacagggcagtc	1920
Qy	1920	aaagcacagaacaaacogtctatggcgtttaatgggggtgaaagtttacttaataataatgcagaaa	1979
Db	1921	aaagcacagaacaaacogtctatggcgtttaatgggggtgaaagtttacttaataataatgcagaaa	1980
Qy	1980	caacagcagaatcgccactactcgtattaccagagataaaattggccttgcctcgagatg	2039
Db	1981	caacagcagaatcgccactactcgtattaccagagataaaattggccttgcctcgagatg	2040
Qy	2040	gtgatgttgataaaaacagaccataatttgataaaaaacaacttaaaagtggttagtg	2099
Db	2041	gtgatgttgataaaaacagaccataatttgataaaaaacaacttaaaagtggttagtg	2100
Qy	2100	ttgcaattaccatagacaaatggcatgtatgcagggttaataaaaaagatcagtaattctgccaa	2159
Db	2101	ttgcaattaccatagacaaatggcatgtatgcagggttaataaaaaagatcagtaattctgccaa	2160
Qy	2160	aaggtagcagtgctaaacgatgcggtttaccatctgaacagctcaaaagccgccaagcctactt	2219
Db	2161	aaggtagcagtgctaaacgatgcggtttaccatctgaacagctcaaaagccgccaagcctactt	2220
Qy	2220	taaacgagggcgtgggcatacagttgcacacctactgaaataatcagttgatgtcaagagtg	2279
Db	2221	taaacgagggcgtgggcatacagttgcacacctactgaaataatcagttgatgtcaagagtg	2280

Qy	2280	gcaatgttaccgcccacattacaacattggtgcgtaaaacaccacgagctttaacagtgatg	2339
Db	2281	gcaatgttaccgcccacattacaacattggtgcgtaaaacaccacgagctttaacagtgatg	2340
Qy	2340	gcactagtataaatttaagttaagggttagtgggtacgaacaaatagcttagttaccgcgcg	2399
Db	2341	gcactagtataaatttagtgggttaagggttagtgggtacgaacaaatagcttagttaccgcgcg	2400
Qy	2400	aacatttggcaagctactataatgaaagtcaatcgcgaacggctgacagtgtctacaagaact	2459
Db	2401	aacatttggcaagctactataatgaaagtcaatcgcgaacggctgacagtgtctacaagaact	2460
Qy	2460	ttaacgttaaagaagaagacgagtgtacgcgcgaacggcttagccagtggtcctaagaatacga	2519
Db	2461	ttaacgttaaagaagaagacgagtgtacgcgcgaacggcttagccagtggtcctaagaatacga	2520
Qy	2520	caaaaaatgcgcgcgacgtacgacatcttaaaactcaaaagttaaaaaacggtctaacgggttg	2579
Db	2521	caaaaaatgcgcgcgacgtacgacatcttaaaactcaaaagttaaaaaacggtctaacgggttg	2580
Qy	2580	ctaccaaaaaagatggttaacggtttacccttgggcttagccaagatagcgggtctgaccatttg	2639
Db	2581	ctaccaaaaaagatggttaacggtttacccttgggcttagccaagatagcgggtctgaccatttg	2640
Qy	2640	gcaaaacacccctaaacacacgctgcttgactgtgttaaagataccacgaacaaatccaag	2699
Db	2641	gcaaaacacccctaaacacacgctgcttgactgtgttaaagataccacgaacaaatccaag	2700
Qy	2700	tcggtgttaatggcatttaaaatttactaaatgtgaaatggtagtaataccaggttaactggcattg	2759
Db	2701	tcggtgttaatggcatttaaaatttactaaatgtgaaatggtagtaataccaggttaactggcattg	2760
Qy	2760	caataacgctgcgattacagagataaaatggcttcttgcttggttctcagttggtgcagttg	2819
Db	2761	caataacgctgcgattacagagataaaatggcttcttgcttggttctcagttggtgcagttg	2820
Qy	2820	atacaacaacaccttacttgcatacagacaagctacaagttggcgaatgttaagattacca	2879
Db	2821	atacaacaacaccttacttgcatacagacaagctacaagttggcgaatgttaagattacca	2880
Qy	2880	acactggcattaaacgcagggtgggttaaagccatcacaggggtgtcccaacactgccttagca	2939
Db	2881	acactggcattaaacgcagggtgggttaaagccatcacaggggtgtcccaacactgccttagca	2940
Qy	2940	ttgcgcgtcaaaatgtagcgcgaacatagaaactgggcgaatacaactccaagacaagaacaat	2999
Db	2941	ttgcgcgtcaaaatgtagcgcgaacatagaaactgggcgaatacaactccaagacaagaacaat	3000
Qy	3000	ccaacgtgcagcgaattaatgatataataatacagggctttaaccttaaaaaataataaca	3059
Db	3001	ccaacgtgcagcgaattaatgatataataatacagggctttaaccttaaaaaataataaca	3060
Qy	3060	accacattgacttgtctccacttagacattgtgactttgcgaatggcaatggcgaacca	3119
Db	3061	accacattgacttgtctccacttagacattgtgactttgcgaatggcaatggcgaacca	3120
Qy	3120	ccgcacagtaaacccatgataccgcttaacaaaaccagtaaaagtggtatatagtatgtaatg	3179
Db	3121	ccgcacagtaaacccatgataccgcttaacaaaaccagtaaaagtggtatatagtatgtaatg	3180
Qy	3180	tgagtgtataaacacattcaatctaacaggcactgatgacaataaaaaacttgcgcgtcaaaa	3239
Db	3181	tgagtgtataaacacattcaatctaacaggcactgatgacaataaaaaacttgcgcgtcaaaa	3240
Qy	3240	ccacaaacttgaaacaaaacagtgtaattggtaatacagcaactaactttaatgttaact	3299
Db	3241	ccacaaacttgaaacaaaacagtgtaattggtaatacagcaactaactttaatgttaact	3300
Qy	3300	ctagtgtatgaagatgccttgtttaacgcgaagaacatcgcgcaaaaatctaacaacccctag	3359
Db	3301	ctagtgtatgaagatgccttgtttaacgcgaagaacatcgcgcaaaaatctaacaacccctag	3360
Qy	3360	ccaaaggaattatcacacaccacaaagccacagacacacgccttacaagaactttaccgttta	3419

Db 3361 ccaaggaaattcacaccacaaagccagacacccgcctacaaacctttaccgtta 3420
Qy 3420 aaaaggtagatgaaaaataataatctgatgacgccacgcacacccgttggtcaaaaga 3479
Db 3421 aaaaggtagatgaaaaataataatctgatgacgccacgcacacccgttggtcaaaaga 3480
Qy 3480 acgcaataatcaagtcacacccctaacactcaaggtgaaacggcttaataattaaaa 3539
Db 3481 acgcaataatcaagtcacacccctaacactcaaggtgaaacggcttaataattaaaa 3540
Qy 3540 ccgacaaaaatggtacgggttaccttttggaattaaacacacaaagcggtcttaaaagccggca 3599
Db 3541 ccgacaaaaatggtacgggttaccttttggaattaaacacacaaagcggtcttaaaagccggca 3600
Qy 3600 aaagcaccttaaacgcacgggtggcttctctatataaaacccacactggtagcgaacaaatcc 3659
Db 3601 aaagcaccttaaacgcacgggtggcttctctatataaaacccacactggtagcgaacaaatcc 3660
Qy 3660 aagtcggtgctgtagggcggtgaagtttccaaggttaataataatggttggtaggtgctg 3719
Db 3661 aagtcggtgctgtagggcggtgaagtttccaaggttaataataatggttggtaggtgctg 3720
Qy 3720 gcatgtatggcacaaactcgcatcaccagagatgaaattggctttactgggactaatggct 3779
Db 3721 gcatgtatggcacaaactcgcatcaccagagatgaaattggctttactgggactaatggct 3780
Qy 3780 cactgtataaagcaaacccacacttaagcaaacgcagcttaacgcaggttggtaaaaga 3839
Db 3781 cactgtataaagcaaacccacacttaagcaaacgcagcttaacgcaggttggtaaaaga 3840
Qy 3840 ttaccaacttcaatcaggtgtagatttccccaaacagccatgatgctgtacagcgcgca 3899
Db 3841 ttaccaacttcaatcaggtgtagatttccccaaacagccatgatgctgtacagcgcgca 3900
Qy 3900 agatttatgatttaaaacgcgaacttgaaacaaaaatcagcagtaactgccaacaaacgac 3959
Db 3901 agatttatgatttaaaacgcgaacttgaaacaaaaatcagcagtaactgccaacaaacgac 3960
Qy 3960 aaaactcattacacgaattctcagtagcagatgaacaggttaataactttacggttagta 4019
Db 3961 aaaactcattacacgaattctcagtagcagatgaacaggttaataactttacggttagta 4020
Qy 4020 accttactocagttatgcacacctcaagacctctgtatgcataccactttgcaggtgaaa 4079
Db 4021 accttactocagttatgcacacctcaagacctctgtatgcataccactttgcaggtgaaa 4080
Qy 4080 acgcatcaccacaaaggttaaaataaaggttgggtgctgtggcgcatcgacacaaacaaag 4139
Db 4081 acgcatcaccacaaaggttaaaataaaggttgggtgctgtggcgcatcgacacaaacaaag 4140
Qy 4140 gcttaacacgccttaagctgacgtgggttaataataatggaagacaggttgcattgaca 4199
Db 4141 gcttaacacgccttaagctgacgtgggttaataataatggaagacaggttgcattgaca 4200
Qy 4200 gccaaaatggtcaaaaatacctacagagactaagcaacactctagttaattgttaccacatg 4259
Db 4201 gccaaaatggtcaaaaatacctacagagactaagcaacactctagttaattgttaccacatg 4260
Qy 4260 ataaagtagcgtacgcacacagacagggcaataataatcaagacagacgacacaaaccc 4319
Db 4261 ataaagtagcgtacgcacacagacagggcaataataatcaagacagacgacacaaaccc 4320
Qy 4320 gtgcgcgcagcatgttgcattgctgaagcgcaggtttaaacttgcaaggaacatggtgaag 4379
Db 4321 gtgcgcgcagcatgttgcattgctgaagcgcaggtttaaacttgcaaggaacatggtgaag 4380
Qy 4380 cgggtgactttgtctccacttatgacacccgtcaactttggcgaatggcaatgccaccacgg 4439
Db 4381 cgggtgactttgtctccacttatgacacccgtcaactttggcgaatggcaatgccaccacgg 4440
Qy 4440 ctaaggtgacctatgatgacacaaagcaaacacagtaaggtgcttatgatcgaatggtg 4499

Db 4441 ctaaggtgacctatgatgacacaaagcaaacacagtaaggtggtctatgatcgaatggtg 4500
Qy 4500 atgatacaacattgaagttaaagataaaaaacttggcgtaaaacacacacattgacca 4559
Db 4501 atgatacaacattgaagttaaagataaaaaacttggcgtaaaacacacacattgacca 4560
Qy 4560 gtactggcacaggtgcttaataaaattgccttaagcaaatcaagctactggcgatgcgcttg 4619
Db 4561 gtactggcacaggtgcttaataaaattgccttaagcaaatcaagctactggcgatgcgcttg 4620
Qy 4620 tcaagggcagtgatgctgtgctcatcaaacacttatctggcgacatccaaactgccca 4679
Db 4621 tcaagggcagtgatgctgtgctcatcaaacacttatctggcgacatccaaactgccca 4680
Qy 4680 aaggggcaagccaaagcaaacactcagcaggtatgtggaatggaatgaaggttca 4739
Db 4681 aaggggcaagccaaagcaaacactcagcaggtatgtggaatggaatgaaggttca 4740
Qy 4740 tctatgacagtaccgataaacaagtactatcaagccaaaaatgtaggacagttgataaaaa 4799
Db 4741 tctatgacagtaccgataaacaagtactatcaagccaaaaatgtaggacagttgataaaaa 4800
Qy 4800 ccaagaagttggccaaagcaaacctggtgccaaagcccaaacccagatggcacattgg 4859
Db 4801 ccaagaagttggccaaagcaaacctggtgccaaagcccaaacccagatggcacattgg 4860
Qy 4860 ctcaaatgaaatgtcaaatcagtcattcaacaagaacaaagtaaatgatacctataaaagc 4919
Db 4861 ctcaaatgaaatgtcaaatcagtcattcaacaagaacaaagtaaatgatacctataaaagc 4920
Qy 4920 aaggcatcaatgaagacaacgcctttgttaaagacttgaaaaagcgccttctctataaca 4979
Db 4921 aaggcatcaatgaagacaacgcctttgttaaagacttgaaaaagcgccttctctataaca 4980
Qy 4980 aaacaaaaaacgcgcagtaactggtgatttaaatgcccgttggccaaacacccgctga 5039
Db 4981 aaacaaaaaacgcgcagtaactggtgatttaaatgcccgttggccaaacacccgctga 5040
Qy 5040 cctttgcaggggatacagggcaaacgcgttaaaactggcgagactttgaccatcaaaag 5099
Db 5041 cctttgcaggggatacagggcaaacgcgttaaaactggcgagactttgaccatcaaaag 5100
Qy 5100 gtgggcaaacagacaccaataagctaacccgataataacatcggtgtgtgtagcaggtactg 5159
Db 5101 gtgggcaaacagacaccaataagctaacccgataataacatcggtgtgtgtagcaggtactg 5160
Qy 5160 atggtctcactgtcaaaccttgcacaaagccttaacaaatcttaacagcgttaatgcaggtg 5219
Db 5161 atggtctcactgtcaaaccttgcacaaagccttaacaaatcttaacagcgttaatgcaggtg 5220
Qy 5220 gcacaaaaatgtgacaaagcggtgtctttttagactcaagcggtcaagcccaaaagcaa 5279
Db 5221 gcacaaaaatgtgacaaagcggtgtctttttagactcaagcggtcaagcccaaaagcaa 5280
Qy 5280 acacccctgtctgaagtgcacaaatggcgctggacccctgggtggcaaggtcatcagtaatg 5339
Db 5281 acacccctgtctgaagtgcacaaatggcgctggacccctgggtggcaaggtcatcagtaatg 5340
Qy 5340 gcaagggcacaaaaagataccgcagctgccaatgtacacagttacaacagttacaacagcaact 5399
Db 5341 gcaagggcacaaaaagataccgcagctgccaatgtacacagttacaacagttacaacagcaact 5400
Qy 5400 tgttgggtcttggtaattgctggttaataacgcgtgacggcaatcaggtataaacattgccc 5459
Db 5401 tgttgggtcttggtaattgctggttaataacgcgtgacggcaatcaggtataaacattgccc 5460
Qy 5460 acataaaaaagaccccaaatcaggttcatcatctaaacccgcactgcatcaaaagcagca 5519
Db 5461 acataaaaaagaccccaaatcaggttcatcatctaaacccgcactgcatcaaaagcagca 5520
Qy 5520 cgggtacttggcggttaaggttaataacgataccgaaaaacttggcactgtgtgtatacaag 5579
Db 5521 cgggtacttggcggttaaggttaataacgataccgaaaaacttggcactgtgtgtatacaag 5580

QY	5580	tgggcgtggataaagacgcgaacgcgtaacgcgcgtatttaagcaatgtttgggtcaaaccc	5639
Db	5581	tgggcgtggataaagacgcgcgaacgcgtaacgcgcgtatttaagcaatgtttgggtcaaaccc	5640
QY	5640	aaaaagatggcagcaaaaaagccctgctcgccacttataagccgcgcgcaggtcgacaccaact	5699
Db	5641	aaaaagatggcagcaaaaaagccctgctcgccacttataagccgcgcgcaggtcgacaccaact	5700
QY	5700	attgacaaacaaaccccgagaaagcattgacagataaaatgaaacaaaggtatcgcgtctct	5759
Db	5701	atttgacaaacaaaccccgagaaagcattgacagataaaatgaaacaaaggtatcgcgtctct	5760
QY	5760	tccatgtcaacgataggccaatcaagagcctgtgtgtacaaagggcgtaaagcgcatgtactcaa	5819
Db	5761	tccatgtcaacgataggccaatcaagagcctgtgtgtacaaagggcgtaaagcgcatgtactcaa	5820
QY	5820	gtgcctaaagcaaacactcagtgcgataggtttccagggccaaagcgagatggtaagccg	5879
Db	5821	gtgcctaaagcaaacactcagtgcgataggtttccagggccaaagcgagatggtaagccg	5880
QY	5880	ccgtttgccataggcagacaaacccaaagcaggcaaccatccatcgccatcggtgataacg	5939
Db	5881	ccgtttgccataggcagacaaacccaaagcaggcaaccatccatcgccatcggtgataacg	5940
QY	5940	cacaagccacggggcgatcaattccatcgccatcggtacaggaatgttggtcagcaggttaagc	5999
Db	5941	cacaagccacggggcgatcaattccatcgccatcggtacaggaatgttggtcagcaggttaagc	6000
QY	6000	actctgtgtccatcggcagcccaagcactgttaagcgtgataaacagttcacagtttggtgta	6059
Db	6001	actctgtgtccatcggcagcccaagcactgttaagcgtgataaacagttcacagtttggtgta	6060
QY	6060	atacaaacagtttacccgatgccatcaaacccgatgtgtttggtgtgggcataaacatca	6119
Db	6061	atacaaacagtttacccgatgccatcaaacccgatgtgtttggtgtgggcataaacatca	6120
QY	6120	ccgtgaccgaagaactcgggtgctcttaggtgtcaaacctctgccatcagtcaggcacac	6179
Db	6121	ccgtgaccgaagaagaactcgggtgctcttaggtgtcaaacctctgccatcagtcaggcacac	6180
QY	6180	acgcagggcacaaagcccaaaaaatctgacgcgcacagcagggtacaaccaccacagcaggtg	6239
Db	6181	acgcagggcacaaagcccaaaaaatctgacgcgcacagcagggtacaaccaccacagcaggtg	6240
QY	6240	caacccgtacggttaaagccttctgctggacaaacggcggttggtgcggtctccgtgggtg	6299
Db	6241	caacccgtacggttaaagccttctgctggacaaacggcggttggtgcggtctccgtgggtg	6300
QY	6300	cctcagggtgctgaacgcctgatcccaaaaatgtggcagcaggttgaggtcagtcgaccacga	6359
Db	6301	cctcagggtgctgaacgcctgatcccaaaaatgtggcagcaggttgaggtcagtcgaccacga	6360
QY	6360	ccgatcgggtcaatgttagcaggttataaaagccacccaaaagcattgtccaaacgaaccca	6419
Db	6361	ccgatcgggtcaatgttagcaggttataaaagccacccaaaagcattgtccaaacgaaccca	6420
QY	6420	atgagcttgaccatcgatatccacaaacagaaataagggccaatgcagggtttctatcag	6479
Db	6421	atgagcttgaccatcgatatccacaaacagaaataagggccaatgcagggtttctatcag	6480
QY	6480	cgatggcgatggcggtccatgccacaagccttacaattcctggcagatccatggtttaccgggg	6539
Db	6481	cgatggcgatggcggtccatgccacaagccttacaattcctggcagatccatggtttaccgggg	6540
QY	6540	gtattgccaccacaaacgggtcaaggtgcggtggcagtgggacgtgcgaagctctcggtata	6599
Db	6541	gtattgccaccacaaacgggtcaaggtgcggtggcagtgggacgtgcgaagctctcggtata	6600
QY	6600	atggtcataatgggtattttaaatcaatgtgttcagccgatcccaagggccatgtatggggcgg	6659
Db	6601	atggtcataatgggtattttaaatcaatgtgttcagccgatcccaagggccatgtatggggcgg	6660

QY	6660	cagttggtgcaggttttccacttttaagccataaatacgcgaagatttttacttaaaaaatcaat	6711
DB	6661		6720
QY	6720	ctcaaccatagttgtatataaaacagatcagcatcagcatctactctgatgctgatgctgttttt	6779
DB	6721		6780
QY	6780	tatcacttaaaacattttaccgctcaagtgatctctttaccaccatgaccaaaatcgccatt	6839
DB	6781		6840
QY	6840	gatcataggttaaacattattgagttaaattttatacaatgtagittgttagatatggttaaaat	6899
DB	6841		6900
QY	6900	gtgcccattgaccacaaaatgaccgattttatcccgaaatcttcgatcatgatccgttga	6959
DB	6901		6960
QY	6960	cctgcaggtcgac	6972
DB	6961	cctgcaggtcgac	6973
<p>RESULT 8</p> <p>US-08-945-567D-1</p> <p>; Sequence 1, Application US/08945567D</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: SASAKI, Ken</p> <p>; APPLICANT: HARKNESS, Robin E.</p> <p>; APPLICANT: LOOMORE, Sheena M.</p> <p>; APPLICANT: CHONG, Pele</p> <p>; APPLICANT: KLEIN, Michel H.</p> <p>; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN C</p> <p>; FILE REFERENCE: 1038-745 MIS</p> <p>; CURRENT APPLICATION NUMBER: US/08/945,567D</p> <p>; CURRENT FILING DATE: 1996-04-29</p> <p>; PRIOR APPLICATION NUMBER: 08/431,718</p> <p>; PRIOR FILING DATE: 1995-05-01</p> <p>; PRIOR APPLICATION NUMBER: 08/478,370</p> <p>; PRIOR FILING DATE: 1995-06-07</p> <p>; PRIOR APPLICATION NUMBER: 08/621,944</p> <p>; PRIOR FILING DATE: 1996-03-26</p> <p>; PRIOR APPLICATION NUMBER: PCT/CA96/00264</p> <p>; PRIOR FILING DATE: 1996-04-29</p> <p>; NUMBER OF SEQ ID NOS: 10</p> <p>; SOFTWARE: PatentIn Ver. 2.1</p> <p>; SEQ ID NO 1</p> <p>; LENGTH: 6973</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Moraxella catarrhalis</p> <p>; FEATURE:</p> <p>; NAME/KEY: CDS</p> <p>; LOCATION: (708)..(6683)</p> <p>US-08-945-567D-1</p>			
<p>Query Match 99.88; Score 6961; DB 13; Length 6973;</p> <p>Best Local Similarity 100.08; Pred. No. 0;</p> <p>Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps</p>			
QY	1	ccatgatatggcagggtgtgctgcctgcctgatgctgcatgacccccattggccc	60
DB	1		60
QY	61	catatctgtacgatttgacatgtgatatttaacatgtgaacatgtttaaacattgttt	120
DB	61		120
QY	121	aatactgttgccatcattaccataaatttagtaacgcatttagtaacgcatttgtaaaat	180

Db 121 aatactgttccatcattaccataaatttagtaaacgcatttagtaaacgcatttgtataaaat 180
Qy 181 cattgcgcccttattgtgtatcatatgaatagaataattattgattgtatctgattattgt 240
Db 181 cattgcgcccttattgtgtatcatatgaatagaataattattgattgtatctgattattgt 240
Qy 241 atcagaatggtgatgctatgatgatgcctacagatgattgatttgggttaatacactctatg 300
Db 241 atcagaatggtgatgctatgatgatgcctacagatgattgatttgggttaatacactctatg 300
Qy 301 atttgatatttttgaactcaactctattgacttaaaacaccataatggttaataatttagca 360
Db 301 atttgatatttttgaactcaactctattgacttaaaacaccataatggttaataatttagca 360
Qy 361 taatggtaggcttttgaataatcacatgcgaattattgttctactgttactaccatgct 420
Db 361 taatggtaggcttttgaataatcacatgcgaattattgttctactgttactaccatgct 420
Qy 421 tgaatgacgatcccaatccacagatcattcaatgtagtattgtttgtatagcaccatttta 480
Db 421 tgaatgacgatcccaatccacagatcattcaatgtagtattgtttgtatagcaccatttta 480
Qy 481 cctcaattatttcaatacaaatgctatgtcagcatgtatcatcttttttaagggttaaacccac 540
Db 481 cctcaattatttcaatacaaatgctatgtcagcatgtatcatcttttttaagggttaaacccac 540
Qy 541 catgaatcacatctataaagtcatctttaaaagccacagggcacattttatggcagtgcc 600
Db 541 catgaatcacatctataaagtcatctttaaaagccacagggcacattttatggcagtgcc 600
Qy 601 agagtacgcaaaatccacacagcac- ggggggggtagctgtgctacagggcaagtggca 659
Db 601 agagtacgcaaaatccacacagcacgggggggtagctgtgctacagggcaagtggca 660
Qy 660 gtgtatcactctgagctttgcccgtattgccgcgtcgtctcctcgtatcggtgcaa 719
Db 661 gtgtatcactctgagctttgcccgtattgccgcgtcgtctcctcgtatcggtgcaa 720
Qy 720 cgctcagtgccagtgcttgcataaaataaaagataccaaacatcgcgaattggtgaac 779
Db 721 cgctcagtgccagtgcttgcataaaataaaagataccaaacatcgcgaattggtgaac 780
Qy 780 aaaaacagcaagcgtcaggcaactccaaaggcgagcgtgacgcagccatgctattg 839
Db 781 aaaaacagcaagcgtcaggcaactccaaaggcgagcgtgacgcagccatgctattg 840
Qy 840 gtgaaaatgttaacgcacagggcggtcaagccatcgccatcggttagtaataaaactg 899
Db 841 gtgaaaatgttaacgcacagggcggtcaagccatcgccatcggttagtaataaaactg 900
Qy 900 tcaatggaacagtttggataagataggtaccgatgctacgggtcaagagtcctacgccca 959
Db 901 tcaatggaacagtttggataagataggtaccgatgctacgggtcaagagtcctacgccca 960
Qy 960 tcggtggtgatgtaaaagcctagtggtgatgcctcgtatggcctcgttagtgatgaacttac 1019
Db 961 tcggtggtgatgtaaaagcctagtggtgatgcctcgtatggcctcgttagtgatgaacttac 1020
Qy 1020 atttgctgtatcagcatggttaactcaaaatccgaaggtactctgtatgaacgatctta 1079
Db 1021 atttgctgtatcagcatggttaactcaaaatccgaaggtactctgtatgaacgatctta 1080
Qy 1080 ttaacgcccattgcagttataaagaatacgaagctcaaaaggataatgatgaataata 1139
Db 1081 ttaacgcccattgcagttataaagaatacgaagctcaaaaggataatgatgaataata 1140
Qy 1140 gacgcacaacccgagcggacacgctcagtgctcagtgaggccatgctatgcacag 1199
Db 1141 gacgcacaacccgagcggacacgctcagtgctcagtgaggccatgctatgcacag 1200
Qy 1200 gtcaattttcaacgctttgtgtacacggggcaacagctcaaaagtgcctattccttgacag 1259
Db 1201 gtcaattttcaacgctttgtgtacacggggcaacagctcaaaagtgcctattccttgacag 1260

Qy 1260 tgggtcttgcgcgcacagccgagggccaaatctacaatcgctatttggcttctgatcaacat 1319
Db 1261 tgggtcttgcgcgcacagccgagggccaaatctacaatcgctatttggcttctgatcaacat 1320
Qy 1320 ctactcgttggagcagatagcccttggtcagagtactcgtcagctacacagggcagta 1379
Db 1321 ctactcgttggagcagatagcccttggtcagagtactcgtcagctacacagggcagta 1380
Qy 1380 ttgccccttaggttcaaggttctgttgcactcagagtataaattcttagaccggcctata 1439
Db 1381 ttgccccttaggttcaaggttctgttgcactcagagtataaattcttagaccggcctata 1440
Qy 1440 caccaaataccagggcactagaccaccccaagtttcaagccacaaataatcacgaagcgggctc 1499
Db 1441 caccaaataccagggcactagaccaccccaagtttcaagccacaaataatcacgaagcgggctc 1500
Qy 1500 cactttccattggttagtaactctatcaaacgtataaaatcatcaatgtcgggtcagaagtta 1559
Db 1501 cactttccattggttagtaactctatcaaacgtataaaatcatcaatgtcgggtcagaagtta 1560
Qy 1560 ataaacccgatcggttcaatgtggtcacagctagaaagcggtgtgaagtgggctaaaggac 1619
Db 1561 ataaacccgatcggttcaatgtggtcacagctagaaagcggtgtgaagtgggctaaaggac 1620
Qy 1620 gtgaatattcttttcaggggtgatgataacagctactgacgttaaaataggtttggataata 1679
Db 1621 gtgaatattcttttcaggggtgatgataacagctactgacgttaaaataggtttggataata 1680
Qy 1680 ctttaactattaaagggtgggtgcagagaccaacgcattaaaccgataataataatcggtg 1739
Db 1681 ctttaactattaaagggtgggtgcagagaccaacgcattaaaccgataataataatcggtg 1740
Qy 1740 taaagagggctgataatagtggtctgaaagttaaactgtctaaaactttaaacaacttta 1799
Db 1741 taaagagggctgataatagtggtctgaaagttaaactgtctaaaactttaaacaacttta 1800
Qy 1800 ctgagggtgaatacaactacataataatgccacaaccacaggttaaggttaggttagtagta 1859
Db 1801 ctgagggtgaatacaactacataataatgccacaaccacaggttaaggttaggttagtagta 1860
Qy 1860 gtactacagctgaattattgagtgatagtttaacctttaccagcccaatacagggcagtc 1919
Db 1861 gtactacagctgaattattgagtgatagtttaacctttaccagcccaatacagggcagtc 1920
Qy 1920 aaagcacaagcaaaacgcgtctatggcgttaatgggtgaaagttaacttaataatgcagaaa 1979
Db 1921 aaagcacaagcaaaacgcgtctatggcgttaatgggtgaaagttaacttaataatgcagaaa 1980
Qy 1980 caacagcagcaatcggcactactcgtattaccagagataaaatggccttctcgagatg 2039
Db 1981 caacagcagcaatcggcactactcgtattaccagagataaaatggccttctcgagatg 2040
Qy 2040 gtgattgttgataaaacaaacacacataatttggataaaaaaacaacttaaaagtggtagtg 2099
Db 2041 gtgattgttgataaaacaaacacacataatttggataaaaaaacaacttaaaagtggtagtg 2100
Qy 2100 ttgcaattaccatagacaatggcattgatgcaggttaataaaagatcagtaactctgccca 2159
Db 2101 ttgcaattaccatagacaatggcattgatgcaggttaataaaagatcagtaactctgccca 2160
Qy 2160 aaggtagcagtgctaaacgatcggtttaccatcgaaacagctcaaaagcccaagcctactt 2219
Db 2161 aaggtagcagtgctaaacgatcggtttaccatcgaaacagctcaaaagcccaagcctactt 2220
Qy 2220 taaacgagggcgctgggcatcagtgctcacactactgaaataatcaggttgatgctaaagatg 2279
Db 2221 taaacgagggcgctgggcatcagtgctcacactactgaaataatcaggttgatgctaaagatg 2280
Qy 2280 gcaatgtttaccgcccccaacttacaacattgcgttgaaacacccagcgcttaacagtgatg 2339
Db 2281 gcaatgtttaccgcccccaacttacaacattgcgttgaaacacccagcgcttaacagtgatg 2340

Db 4501 atgatacaacaccattgaagttataagataaaaaaacttggcgttaaaaaacccaccacattgacca 4560
Qy 4560 gtactggcacaggctgctaataaatttgccttaagcaaatcaagactactggcgactgtg 4619
Db 4561 gtactggcacaggctgctaataaatttgccttaagcaaatcaagactactggcgactgtg 4620
Qy 4620 tcaaggccagtgatattgcttactatataacaccttatcttgcgacatccaaactggcca 4679
Db 4621 tcaaggccagtgatattgcttactatataacaccttatcttgcgacatccaaactggcca 4680
Qy 4680 aaggggcaagccaaagcgaacaaactcagcagctatgtggatgcttgaaggcaataaaggctca 4739
Db 4681 aaggggcaagccaaagcgaacaaactcagcagctatgtggatgcttgaaggcaataaaggctca 4740
Qy 4740 tctatgacagtaccgataaacaagtactatcaagccaaaaatgtagggcacagttgataaaaa 4799
Db 4741 tctatgacagtaccgataaacaagtactatcaagccaaaaatgtagggcacagttgataaaaa 4800
Qy 4800 ccaagaagaattgccaagacaacactgtgcgccaaagcccaaacccacagatggcacattgg 4859
Db 4801 ccaagaagaattgccaagacaacactgtgcgccaaagcccaaacccacagatggcacattgg 4860
Qy 4860 ctcaaatgaattgcaaatcagctatcaacaagaacaaagtaaatgtagtcccaataaaaaagc 4919
Db 4861 ctcaaatgaattgcaaatcagctatcaacaagaacaaagtaaatgtagtcccaataaaaaagc 4920
Qy 4920 aaggcatcaatggaagacaacaccttgttaaaggacttgaaaaagccgcttctgataaaca 4979
Db 4921 aaggcatcaatggaagacaacaccttgttaaaggacttgaaaaagccgcttctgataaaca 4980
Qy 4980 aaacaaaaaacgcgcgagtaactgtgggtgatttaaatgccgttgcccaaacacacgcgtga 5039
Db 4981 aaacaaaaaacgcgcgagtaactgtgggtgatttaaatgccgttgcccaaacacacgcgtga 5040
Qy 5040 cctttgcaagggtatcacaggcaacacgcgttaaaaaactgggcgaactttgaccatcaaaag 5099
Db 5041 cctttgcaagggtatcacaggcaacacgcgttaaaaaactgggcgaactttgaccatcaaaag 5100
Qy 5100 gtgggcaaacagacaccaataagctaacccgataataacatcggtgtggtagcaggtactg 5159
Db 5101 gtgggcaaacagacaccaataagctaacccgataataacatcggtgtggtagcaggtactg 5160
Qy 5160 atggtctcaactgtcaaaacttgcgaagaccttaacaaactttaaagcgttaaatgcaggtg 5219
Db 5161 atggtctcaactgtcaaaacttgcgaagaccttaacaaactttaaagcgttaaatgcaggtg 5220
Qy 5220 gcacaaaaattgatacaaaaagcgtgtctttttagactcaagcgttcaagcccaaaagcaa 5279
Db 5221 gcacaaaaattgatacaaaaagcgtgtctttttagactcaagcgttcaagcccaaaagcaa 5280
Qy 5280 acacccctgtgctgaagtgcgaatgggctggacctgggtggcgaaggtcatcagtaagtgg 5339
Db 5281 acacccctgtgctgaagtgcgaatgggctggacctgggtggcgaaggtcatcagtaagtgg 5340
Qy 5340 gcaaggcacaataagataccgacgctgccaatgtacaacagtttaaacgaagtacgcaact 5399
Db 5341 gcaaggcacaataagataccgacgctgccaatgtacaacagtttaaacgaagtacgcaact 5400
Qy 5400 tgttgggtcttgggttaattgctgtaataacgctgacggccaatcagggtaaacattgccc 5459
Db 5401 tgttgggtcttgggttaattgctgtaataacgctgacggccaatcagggtaaacattgccc 5460
Qy 5460 acataaaaaagaccccaaatcaggttcatcatctaaccgcactgtcatcaaaagcaggca 5519
Db 5461 acataaaaaagaccccaaatcaggttcatcatctaaccgcactgtcatcaaaagcaggca 5520
Qy 5520 cgggtacttggcgttaagggttaataacgataccggaataaacttgcacgttgggtatacaag 5579
Db 5521 cgggtacttggcgttaagggttaataacgataccggaataaacttgcacgttgggtatacaag 5580
Qy 5580 tgggcgtggataaagacggaacgcctaacgcgatttaagcaaatgtttgggtcaaaacccc 5639
Db 5581 tgggcgtggataaagacggaacgcctaacgcgatttaagcaaatgtttgggtcaaaacccc 5640

Qy 5640 aaaaagatggcagcaaaaaagccctgtcgcacctataaacgccgcaggttcagaccaact 5699
Db 5641 aaaaagatggcagcaaaaaagccctgtcgcacctataaacgccgcaggttcagaccaact 5700
Qy 5700 atttgaccaacaacccccgcagaagccattgacagaataaatagaacaaggtatccgcttct 5759
Db 5701 atttgaccaacaacccccgcagaagccattgacagaataaatagaacaaggtatccgcttct 5760
Qy 5760 tccatgtcaacgatggcaatcaagagcctgtggtaacaaggcgttaacggcatttgactcaa 5819
Db 5761 tccatgtcaacgatggcaatcaagagcctgtggtaacaaggcgttaacggcatttgactcaa 5820
Qy 5820 gtgcctcaggcaagcaactcagtgggcgtataggtttccaggcccaagcagatgggtgaagcc 5879
Db 5821 gtgcctcaggcaagcaactcagtgggcgtataggtttccaggcccaagcagatgggtgaagcc 5880
Qy 5880 ccgttgcataaggcagacaaccccccaagcaggaacaaatccatcccatcggtgataaag 5939
Db 5881 ccgttgcataaggcagacaaccccccaagcaggaacaaatccatcccatcggtgataaag 5940
Qy 5940 cacaagccacagggcgatcaatccatccgcctacgcgttacaggccaatgtgtacgagtaagc 5999
Db 5941 cacaagccacagggcgatcaatccatccgcctacgcgttacaggccaatgtgtacgagtaagc 6000
Qy 6000 actctggtgcctacggcgaccacaagcactgttaagcgtgataaacagtttacagttgggta 6059
Db 6001 actctggtgcctacggcgaccacaagcactgttaagcgtgataaacagtttacagttgggta 6060
Qy 6060 ataacaacacagttaccgatgccactcaaacogatactgtttggtgggcaataacatca 6119
Db 6061 ataacaacacagttaccgatgccactcaaacogatactgtttggtgggcaataacatca 6120
Qy 6120 ccgtgacagaaagtaactcggttgccttaggttcaaaactctgccatcagtcagcagcaac 6179
Db 6121 ccgtgacagaaagtaactcggttgccttaggttcaaaactctgccatcagtcagcagcaac 6180
Qy 6180 acgagggcacacaagccaaaaatctgcgcgcacagcaggttacaccaccacacagcaggtg 6239
Db 6181 acgagggcacacaagccaaaaatctgcgcgcacagcaggttacaccaccacacagcaggtg 6240
Qy 6240 caacccgttacaggttaaaagccttgcgtgacaaaacggcgttgcgtgcgttccgtgcgtg 6299
Db 6241 caacccgttacaggttaaaagccttgcgtgacaaaacggcgttgcgtgcgttccgtgcgtg 6300
Qy 6300 cctcaggtgcctgaacgcgctatccaaaatgtggcagcaggtgagctcagtcgacacagca 6359
Db 6301 cctcaggtgcctgaacgcgctatccaaaatgtggcagcaggttgcgtgcgttccacacagca 6360
Qy 6360 ccgatgcggttcaatgggtgagccagttgtacaagcccaaccccaagcattgccaaacgaacca 6419
Db 6361 ccgatgcggttcaatgggtgagccagttgtacaagcccaaccccaagcattgccaaacgaacca 6420
Qy 6420 atgagcttgacacatcgtatccacaaaacgaaaaataagggccaatgcagggatttcatcag 6479
Db 6421 atgagcttgacacatcgtatccacaaaacgaaaaataagggccaatgcagggatttcatcag 6480
Qy 6480 cgatggcgatggcgtccatgcacaagcctacatcctggcagatcccatgggttaccgggg 6539
Db 6481 cgatggcgatggcgtccatgcacaagcctacatcctggcagatcccatgggttaccgggg 6540
Qy 6540 gtattggcaccaccaaacgcgtcaaggtgcgttggcaggtgggactgtcgaagcgttcgggata 6599
Db 6541 gtattggcaccaccaaacgcgtcaaggtgcgttggcaggtgggactgtcgaagcgttcgggata 6600
Qy 6600 atggttcaatgggttataaatacaatgggttcagccgataaccaagccatgttagggcggg 6659
Db 6601 atggttcaatgggttataaatacaatgggttcagccgataaccaagccatgttagggcggg 6660
Qy 6660 cagttggtgcaggttttcaactttaagccataaatacgaagattttacttaaaatacaat 6719
Db 6661 cagttggtgcaggttttcaactttaagccataaatacgaagattttacttaaaatacaat 6720

Qy	6720	ctcaccatagttgtataaaacagcatcagcatcattactgactgctgattgtttt	6779
Db	6721	ctcaccatagttgtataaaacagcatcagcatcattactgactgctgattgtttt	6780
Qy	6780	tatcacttaaacattttacgcgtccaagtgtctctctccaccatgaccaaactcgccatt	6839
Db	6781	tatcacttaaacattttacgcgtccaagtgtctctctccaccatgaccaaactcgccatt	6840
Qy	6840	gatcataggtaaaacttttagtgaataattttatcaatgtagtgttagatatgtttaaat	6899
Db	6841	gatcataggtaaaacttttagtgaataattttatcaatgtagtgttagatatgtttaaat	6900
Qy	6900	tgtgcattgacaaaataaacgagattttatccgaaaaatttctgattatgatccgttga	6959
Db	6901	tgtgcattgacaaaataaacgagattttatccgaaaaatttctgattatgatccgttga	6960
Qy	6960	cctgcaggtcgac	6972
Db	6961	cctgcaggtcgac	6973

RESULT 9
 US-09-361-619-1
 ; Sequence 1, Application US/09361619
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Sasaki, Ken
 ; APPLICANT: Yang, Yan ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
 ; TITLE OF INVENTION: PROTEIN OF MORAXELLA
 ; FILE REFERENCE: 1038-921MIS:jb
 ; CURRENT APPLICATION NUMBER: US/09/361,619
 ; CURRENT FILING DATE: 1999-07-27
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 6973
 ; TYPE: DNA
 ; ORGANISM: Moraxella catarrhalis
 US-09-361-619-1

Qy	1	ccatgatatggcgaggtgctgcgcgcgcgtatgatggcgatgacacccatttgcgc	60
Db	1	ccatgatatggcgaggtgctgcgcgcgcgtatgatggcgatgacacccatttgcgc	60
Qy	61	catatctgtacgatttgacatgtatgattaaacatgtgacatgatttaacattgttt	120
Db	61	catatctgtacgatttgacatgtatgattaaacatgtgacatgatttaacattgttt	120
Qy	121	aatactgttgcattccattaccataatttagtaaacgatttagtaaacgatttgaataaat	180
Db	121	aatactgttgcattccattaccataatttagtaaacgatttagtaaacgatttgaataaat	180
Qy	181	cattgcgcccctttatgtgtatcatgaatgaataattatgattgtatctgattattgt	240
Db	181	cattgcgcccctttatgtgtatcatgaatgaataattatgattgtatctgattattgt	240
Qy	241	atcagaatggtgtgcttatatgatgatgctcacgagttgatttgggttaactactctatg	300
Db	241	atcagaatggtgtgcttatatgatgatgctcacgagttgatttgggttaactactctatg	300
Qy	301	atttgatatatttgaactaaactatttgacttaaacacacatggtttaatttagca	360
Db	301	atttgatatatttgaactaaactatttgacttaaacacacatggtttaatttagca	360
Qy	361	taatggtgaggtttttgaaaaatcacatgcgaatttcttactgttactaccattgt	420
Db	361	taatggtgaggtttttgaaaaatcacatgcgaatttcttactgttactaccattgt	420

Query Match 99.8%; Score 6961; DB 17; Length 6973;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6972; Conservative 0; Mismatches 0; Indels 1; Gaps 1

QY 1500 caatttccatttggttagtaactcta tcaaacgttaaaatcaatcaatgcaggtgttta 1559
DB 1501 caatttccatttggttagtaactcta tcaaacgttaaaatcaatcaatgcaggtgttta 1560
QY 1560 ataaacacgagtcggtcaatggtgcacagctagaagcgggtggtgaagtggtctaaaggagc 1619
DB 1561 aaaaacacgagtcggtcaatggtgcacagctagaagcgggtggtgaagtggtctaaaggagc 1620
QY 1620 gtagaattactttcaggtggtatgaataacagctactgaagcgaataatggtttggataata 1679
DB 1621 gtagaattactttcaggtggtatgaataacagctactgaagcgaataatggtttggataata 1680
QY 1680 ctttaactattaaaggtggtgcagagcacaacgcattaccgcgataataatgcggtgtgg 1739
DB 1681 ctttaactattaaaggtggtgcagagcacaacgcattaccgcgataataatgcggtgtgg 1740
QY 1740 taaaagaggtgcataatagttggttgaaagttaaaacttgctaaaactttaaacatactta 1799
DB 1741 taaaagaggtgcataatagttggttgaaagttaaaacttgctaaaactttaaacatactta 1800
QY 1800 ctgaggtgaatacaactacattaaatgccacaacacagcttaaggtaggtagtagtagta 1859
DB 1801 ctgaggtgaatacaactacattaaatgccacaacacagcttaaggtaggtagtagtagta 1860
QY 1860 gtactacaggtgaattattgagtagtagtttaacctttaccagcccaatacaggcagtc 1919
DB 1861 gtactacaggtgaattattgagtagtagtttaacctttaccagcccaatacaggcagtc 1920
QY 1920 aaagcacaagaacaaacgctctatgcgttaattggtgaagtttacttaataatgcagaaa 1979
DB 1921 aaagcacaagaacaaacgctctatgcgttaattggtgaagtttacttaataatgcagaaa 1980
QY 1980 caacagcagcaatcgccactactcgtattaccagagataaaattggtttgctgcagatg 2039
DB 1981 caacagcagcaatcgccactactcgtattaccagagataaaattggtttgctgcagatg 2040
QY 2040 gtgagtgtgatgaataaagaacacacattttggataaaacacacattaaagtgggtagtg 2099
DB 2041 gtgagtgtgatgaataaagaacacacattttggataaaacacacattaaagtgggtagtg 2100
QY 2100 ttgcaattaccatagacaaatgcatgagtgagtgagtgagtgagtgagtgagtgagtgag 2159
DB 2101 ttgcaattaccatagacaaatgcatgagtgagtgagtgagtgagtgagtgagtgagtgag 2160
QY 2160 aaggtagcagtgtaacgagtcggttaccatgcagacgctcaaacgcccagcctacttt 2219
DB 2161 aaggtagcagtgtaacgagtcggttaccatgcagacgctcaaacgcccagcctacttt 2220
QY 2220 taaacgagcgtggttcagtgatgcacacactactgaaataatcagttgagtgtaagtg 2279
DB 2221 taaacgagcgtggttcagtgatgcacacactactgaaataatcagttgagtgtaagtg 2280
QY 2280 gcaattgtaccgcccacacttaacattgggtgtaaaacacacgagcttaacagtgatg 2339
DB 2281 gcaattgtaccgcccacacttaacattgggtgtaaaacacacgagcttaacagtgatg 2340
QY 2340 gcaactgataaaattagtgtaagggtagtggtacgaacacacacagcttagttaccgcg 2399
DB 2341 gcaactgataaaattagtgtaagggtagtggtacgaacacacacagcttagttaccgcg 2400
QY 2400 aacatttggcagcgtatcattaaatgaagtgaatgcgaacggtgacagtgctctacaagct 2459
DB 2401 aacatttggcagcgtatcattaaatgaagtgaatgcgaacggtgacagtgctctacaagct 2460
QY 2460 ttaccgttaaaagaagacgagtgatgacgccaacgctataccggtggtgataagatacga 2519
DB 2461 ttaccgttaaaagaagacgagtgatgacgccaacgctataccggtggtgataagatacga 2520
QY 2520 caaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctaacggttg 2579
DB 2521 caaaaaatgcggcgagtcagcatcttaaaactcaaaaggttaaaacggtctaacggttg 2580

QY 2580 ctacaaaaaagatggttagctgttaactcttggtcttagccaagatagcgtgctgacacatg 2639
DB 2581 ctacaaaaaagatggttagctgttaactcttggtcttagccaagatagcgtgctgacacatg 2640
QY 2640 gcaaaagcaccctaaacaacagatgctgtgactgttaaaatgatacacaacacacacacag 2699
DB 2641 gcaaaagcaccctaaacaacagatgctgtgactgttaaaatgatacacaacacacacacag 2700
QY 2700 tcggtgttaattggcattaaatttactaaattgtaattggttagtaattccaggtactggcattg 2759
DB 2701 tcggtgttaattggcattaaatttactaaattgtaattggttagtaattccaggtactggcattg 2760
QY 2760 caaataccgctgcgattaccagagataaaattggctttgctggttctggtgagtgagtg 2819
DB 2761 caaataccgctgcgattaccagagataaaattggctttgctggttctggtgagtgagtg 2820
QY 2820 atacaaaaaaaccttatctgtcatcaagacacagctacaagttgcaatgttaagattacca 2879
DB 2821 atacaaaaaaaccttatctgtcatcaagacacagctacaagttgcaatgttaagattacca 2880
QY 2880 acactggcattaacgcaggtggttaaaagccatcacagggctgtccccaacactgcctagca 2939
DB 2881 acactggcattaacgcaggtggttaaaagccatcacagggctgtccccaacactgcctagca 2940
QY 2940 ttgcccagatcaaaagtagcgcacacatagaactgggcaatacaatccaaagacacacaaat 2999
DB 2941 ttgcccagatcaaaagtagcgcacacatagaactgggcaatacaatccaaagacacacaaat 3000
QY 3000 ccaacgctgcagcattaaatgatatataatatacaggttttaaccttaaaaaataataca 3059
DB 3001 ccaacgctgcagcattaaatgatatataatatacaggttttaaccttaaaaaataataca 3060
QY 3060 acccattgactttgtctccacttgatgactgtgactttgccaatggcgaatggcaccaca 3119
DB 3061 acccattgactttgtctccacttgatgactgtgactttgccaatggcgaatggcaccaca 3120
QY 3120 ccgcccagatcaacccatgataccgctaaacaaacacagtaaaagtgttatgtatgtgaatg 3179
DB 3121 ccgcccagatcaacccatgataccgctaaacaaacacagtaaaagtgttatgtatgtgaatg 3180
QY 3180 ttgattgatacaaccattctatcaacaggcacatgacacataaaaaaacttggcgtcaaaa 3239
DB 3181 ttgattgatacaaccattctatcaacaggcacatgacacataaaaaaacttggcgtcaaaa 3240
QY 3240 ccacaaactgaacaaacaaagtgttaattggttaatacagcaactaaacttttaattgtaact 3299
DB 3241 ccacaaactgaacaaacaaagtgttaattggttaatacagcaactaaacttttaattgtaact 3300
QY 3300 ctgattgataagatgccttgttaacgccaagacatcgccgaataatctaaacacacctag 3359
DB 3301 ctgattgataagatgccttgttaacgccaagacatcgccgaataatctaaacacacctag 3360
QY 3360 ccaagaaattcacaccacaaagccacagcagacacccctacaaaactttacacgtta 3419
DB 3361 ccaagaaattcacaccacaaagccacagcagacacccctacaaaactttacacgtta 3420
QY 3420 aaaaagtgaatgaaaaataataatgctgagcgaacccacacacacacacacacacacacac 3479
DB 3421 aaaaagtgaatgaaaaataataatgctgagcgaacccacacacacacacacacacacacac 3480
QY 3480 acgcaataatacgaatcaacacccctaaactcaactcaaaaggtgaaaacggtcttaataata 3539
DB 3481 acgcaataatacgaatcaacacccctaaactcaactcaaaaggtgaaaacggtcttaataata 3540
QY 3540 ccgcaaaaatgtagcgtttactttggcattaaacacacacacacacacacacacacacacac 3599
DB 3541 ccgcaaaaatgtagcgtttactttggcattaaacacacacacacacacacacacacacacac 3600
QY 3600 aaagcaccctaaacacggtggtctgtctattaaaaacccacacacacacacacacacacacac 3659
DB 3601 aaagcaccctaaacacggtggtctgtctattaaaaacccacacacacacacacacacacacac 3660
QY 3660 aagtggtgctgagtggtgaggtttgccaaggttaataataatggtgtgtgtgagtgctg 3719

||||| aegtcggtgctgagtcgtaagttgcgaaggttaataataatggtgtgtagtgctg 3720 Db
3720 gcatatgagccaaactcgcattaccagagatgaaattgctttactggactaatggct 3779 Qy
3721 gcatatgagccaaactcgcattaccagagatgaaattgctttactggactaatggct 3780 Db
3780 cacttgataaaagcaaacccacttaagcaagacggttaacgagtggttaaaaga 3839 Qy
3781 cacttgataaaagcaaacccacttaagcaagacggttaacgagtggttaaaaga 3840 Db
3840 ttaccaacttcaatcaggttgagattgccccaaacgcccattgctgtgacagggcgca 3899 Qy
3841 ttaccaacttcaatcaggttgagattgccccaaacgcccattgctgtgacagggcgca 3900 Db
3900 agatttatgatttaaaacgaacttgaaaacaaatcagcagtagtgcgaacacgac 3959 Qy
3901 agatttatgatttaaaacgaacttgaaaacaaatcagcagtagtgcgaacacgac 3960 Db
3960 aaactcattacgaattctcagtagcagatgaacaaggttaataactttaagggttagta 4019 Qy
3961 aaactcattacgaattctcagtagcagatgaacaaggttaataactttaagggttagta 4020 Db
4020 acccttactccagttatgacacotcaaacctctctgattcattcacctttgcaagtgaaa 4079 Qy
4021 acccttactccagttatgacacotcaaacctctctgattcattcacctttgcaagtgaaa 4080 Db
4080 acggcattacaccaaaggttaataaaaggttggtgctgctggtgacattgacaccaaag 4139 Qy
4081 acggcattacaccaaaggttaataaaaggttggtgctgctggtgacattgacaccaaag 4140 Db
4140 gcttaacccagcctaagctgacccgtgggttaataaatggcaaggaattgctattgaca 4199 Qy
4141 gcttaacccagcctaagctgacccgtgggttaataaatggcaaggaattgctattgaca 4200 Db
4200 gcaaaaatgctaaaaatcacatcacaggaactgaacacotctagtaattgtaccaatg 4259 Qy
4201 gcaaaaatgctaaaaatcacatcacaggaactgaacacotctagtaattgtaccaatg 4260 Db
4260 ataaagtagcglacgaccacagaaacagggcaataatacaaaagacgaagacaaaacc 4319 Qy
4261 ataaagtagcglacgaccacagaaacagggcaataatacaaaagacgaagacaaaacc 4320 Db
4320 gtgcgcgcagcattgtgtgctgtaagcagcaggtttaaacttgcaagggcaatgggtgaag 4379 Qy
4321 gtgcgcgcagcattgtgtgctgtaagcagcaggtttaaacttgcaagggcaatgggtgaag 4380 Db
4380 cgggtgacttctccacttatgacacccgtcaacttgcgagtggaatgcccacacgg 4439 Qy
4381 cgggtgacttctccacttatgacacccgtcaacttgcgagtggaatgcccacacgg 4440 Db
4440 ctaaggtgacttatgtagacacagcaaaaccagtaaaagtggttctatgtagtcaatgtgg 4499 Qy
4441 ctaaggtgacttatgtagacacagcaaaaccagtaaaagtggttctatgtagtcaatgtgg 4500 Db
4500 atgatacaaccattgaagttaaagataaaaaacttgcgttaaaaaaccacacattgacca 4559 Qy
4501 atgatacaaccattgaagttaaagataaaaaacttgcgttaaaaaaccacacattgacca 4560 Db
4560 gtactgacaggtgcttaataaatttgccttaagcaatcaagctactggcgatgcgcttg 4619 Qy
4561 gtactgacaggtgcttaataaatttgccttaagcaatcaagctactggcgatgcgcttg 4620 Db
4620 tcaaggccagtgatgctgtctatctaaacaccttatctggcgacatccaaactgcca 4679 Qy
4621 tcaaggccagtgatgctgtctatctaaacaccttatctggcgacatccaaactgcca 4680 Db
4680 aaggggcaagccagcaaacactcagcaggtgatggtgctgagtggaataaggtca 4739 Qy
4681 aaggggcaagccagcaaacactcagcaggtgatggtgctgagtggaataaggtca 4740 Db
4740 tctatgacagttaccgataaacaagttactatcaagccaaaatgatggcacaggttgataaa 4799 Qy
|||||

4741 tctatgacagttaccgataaacaagttactatcaagccaaaatgtagtggcacagttgataaaa 4800 Db
4800 ccaaaagagttgcccacaaagacaaactggtcgcccacagcccaaacccagatggcacattgg 4859 Qy
4801 ccaaaagagttgcccacaaagacaaactggtcgcccacagcccaaacccagatggcacattgg 4860 Db
4860 ctcaaatgaattcaaatcagtcatttaacaagaacaagtaaatgatgccaataaaaagc 4919 Qy
4861 ctcaaatgaattcaaatcagtcatttaacaagaacaagtaaatgatgccaataaaaagc 4920 Db
4920 aaggcatcaatgaagacacacgcttctgttaaaggacttgaaaaagcgcgttctgtgataaca 4979 Qy
4921 aaggcatcaatgaagacacacgcttctgttaaaggacttgaaaaagcgcgttctgtgataaca 4980 Db
4980 aaacccaaaacgcccagtaactgtgggtgatttaaatccgttgcgccaaaacacgcgtga 5039 Qy
4981 aaacccaaaacgcccagtaactgtgggtgatttaaatccgttgcgccaaaacacgcgtga 5040 Db
5040 cctttgcagggatatacagggcacacaggttaaaaactggcgagactttgacccatcaaaag 5099 Qy
5041 cctttgcagggatatacagggcacacaggttaaaaactggcgagactttgacccatcaaaag 5100 Db
5100 gtgggcaaacagacacacaaataagctaaacgcgataataacatcggtgtgtgtagcaggtactg 5159 Qy
5101 gtgggcaaacagacacacaaataagctaaacgcgataataacatcggtgtgtgtagcaggtactg 5160 Db
5160 atggttctactgtcaaaactgcccagacactaaacaaatcttaaacacgcttaattgcaaggtg 5219 Qy
5161 atggttctactgtcaaaactgcccagacactaaacaaatcttaaacacgcttaattgcaaggtg 5220 Db
5220 gaacccaaaatgatgacaaaagcgtgtctttgttagactcaagcgttcaagccaaaagcaaa 5279 Qy
5221 gaacccaaaatgatgacaaaagcgtgtctttgttagactcaagcgttcaagccaaaagcaaa 5280 Db
5280 acacccctgtgctaaatgcccattgggctggacgtgggtggcaaggttcacagtaattgtgg 5339 Qy
5281 acacccctgtgctaaatgcccattgggctggacgtgggtggcaaggttcacagtaattgtgg 5340 Db
5340 gcaaaagccaaaagatacagcagctgcccattgacaacagttacaacaggttaaacgaagtacgcaact 5399 Qy
5341 gcaaaagccaaaagatacagcagctgcccattgacaacagttacaacaggttaaacgaagtacgcaact 5400 Db
5400 tgttggttctgtgtaattgctggttaataatgaacgtgacgcaatcaggttaaacattgccc 5459 Qy
5401 tgttggttctgtgtaattgctggttaataatgaacgtgacgcaatcaggttaaacattgccc 5460 Db
5460 acatcaaaaaagccccaaattcaggttcatctcaacgcactgtcatcaaaagcaggca 5519 Qy
5461 acatcaaaaaagccccaaattcaggttcatctcaacgcactgtcatcaaaagcaggca 5520 Db
5520 cgttacttggcgttaaaaggttaataacgataccgaaaaacttgcacttgggtggtatacaag 5579 Qy
5521 cgttacttggcgttaaaaggttaataacgataccgaaaaacttgcacttgggtggtatacaag 5580 Db
5580 tgggctgggttaaaagcggcaacgcgttaacgcgtatttaagcaaatgtttgggttcaaaaccc 5639 Qy
5581 tgggctgggttaaaagcggcaacgcgttaacgcgtatttaagcaaatgtttgggttcaaaaccc 5640 Db
5640 aaaaaagtgagcaaaaaagccccctgctgcacttataacgcgcaggttcagaccaact 5699 Qy
5641 aaaaaagtgagcaaaaaagccccctgctgcacttataacgcgcaggttcagaccaact 5700 Db
5700 atttgacccaaaccccccgacagaagcattgacagaaataatgaacaaggttatccgcttct 5759 Qy
5701 atttgacccaaaccccccgacagaagcattgacagaaataatgaacaaggttatccgcttct 5760 Db
5760 tccatgtcaacagatggcaatcaagagcgtgtgtgataagggcgtaacggttactgactcaa 5819 Qy
5761 tccatgtcaacagatggcaatcaagagcgtgtgtgataagggcgtaacggttactgactcaa 5820 Db
5820 gtgcctcaggaagcaactcagtgagtaggtttccagggccaagcagatgggtgaagccg 5879 Qy
5821 gtgcctcaggaagcaactcagtgagtaggtttccagggccaagcagatgggtgaagccg 5880 Db
|||||

Qy 5880 ccgttgccataggcagacaaaccccaagcaggcaacccaaatccatccatcggtgataaacg 5939
Db 5881 ccgttgccataggcagacaaaccccaagcaggcaacccaaatccatccatcggtgataaacg 5940
Qy 5940 cacaagccacagggcgagataaattccatcgccatcggttacaggcaaatgtgtagcaggttaagc 5999
Db 5941 cacaagccacagggcgagataaattccatcgccatcggttacaggcaaatgtgtagcaggttaagc 6000
Qy 6000 actctgggtgcatcgccagcccaagcactgttaaggctgataaacagttacagttggtggtga 6059
Db 6001 actctgggtgcatcgccagcccaagcactgttaaggctgataaacagttacagttggtggtga 6060
Qy 6060 ataacacacagttaccgatgccactcaaacccgatgtcttgggtggtggcgaataaacatca 6119
Db 6061 ataacacacagttaccgatgccactcaaacccgatgtcttgggtggtggcgaataaacatca 6120
Qy 6120 ccgtgacccaagaataactcggttgcccttagtttcaaacctcgccatcagtcagggcacac 6179
Db 6121 ccgtgacccaagaataactcggttgcccttagtttcaaacctcgccatcagtcagggcacac 6180
Qy 6180 acgcagggcacacagccaacaaatctgacgcacagcaggttacaaccaccacagcaggtg 6239
Db 6181 acgcagggcacacagccaacaaatctgacgcacagcaggttacaaccaccacagcaggtg 6240
Qy 6240 caaccgggtacgggttaaggcttctgagcaaaacggcggttgggtgcggtctccgtgggtg 6299
Db 6241 caaccgggtacgggttaaggcttctgagcaaaacggcggttgggtgcggtctccgtgggtg 6300
Qy 6300 cctcaggtgctgaaccccgatcaacaaaatgtggcagcaggttgaagtcagtcaccaccagca 6359
Db 6301 cctcaggtgctgaaccccgatcaacaaaatgtggcagcaggttgaagtcagtcaccaccagca 6360
Qy 6360 ccgattgcggtcaatggtagcaggttgcacaaagccaccacaaagcattgccaacgcaacca 6419
Db 6361 ccgattgcggtcaatggtagcaggttgcacaaagccaccacaaagcattgccaacgcaacca 6420
Qy 6420 atgagcttgaccatcgatccaccacaaacgaaaaataaggcccaatgcagggatttcacag 6479
Db 6421 atgagcttgaccatcgatccaccacaaacgaaaaataaggcccaatgcagggatttcacag 6480
Qy 6480 cgatggcgatgctccatgcacaaagcctcattctcggcagatccatggttaccgggg 6539
Db 6481 cgatggcgatgctccatgcacaaagcctcattctcggcagatccatggttaccgggg 6540
Qy 6540 gtattgccaccacaaacggtcaaggtgcggtggcagtgaggactgctgaagctgtcggtata 6599
Db 6541 gtattgccaccacaaacggtcaaggtgcggtggcagtgaggactgctgaagctgtcggtata 6600
Qy 6600 atggttcaatgggtatttaaaatcaatggttcagccgataaccgaaggccatgtagggcg 6659
Db 6601 atggttcaatgggtatttaaaatcaatggttcagccgataaccgaaggccatgtagggcg 6660
Qy 6660 cagttggtgcaggttttcaactttaagccataaatacgaagatttacttaaaatacaat 6719
Db 6661 cagttggtgcaggttttcaactttaagccataaatacgaagatttacttaaaatacaat 6720
Qy 6720 ctacacatagttgtataaaacagcatcagcatcagtcataattactgctgtagtgtttt 6779
Db 6721 ctacacatagttgtataaaacagcatcagcatcagtcataattactgctgtagtgtttt 6780
Qy 6780 tatcaacttaaacattttaccgcctcaagtgattctcttcttccaccatgacccaatcgccatt 6839
Db 6781 tatcaacttaaacattttaccgcctcaagtgattctcttcttccaccatgacccaatcgccatt 6840
Qy 6840 gatcataggttaaacatttttagtaaaattttatcaatgattgtttagatagtttaaaat 6899
Db 6841 gatcataggttaaacatttttagtaaaattttatcaatgattgtttagatagtttaaaat 6900
Qy 6900 tgtgcatgtgacaaaataatgaccgatttatcccgaaaatttctgatttatgatccgttga 6959
Db 6901 tgtgcatgtgacaaaataatgaccgatttatcccgaaaatttctgatttatgatccgttga 6960

Qy 6960 cctgcaggtcgac 6972
Db 6961 cctgcaggtcgac 6973
RESULT 10
US-08-431-718-1
; Sequence 1, Application US/08431718
; GENERAL INFORMATION:
; APPLICANT: Sasaki, Ken
; APPLICANT: Harkness, Robin E
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,718
; FILING DATE: 01-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-429
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-431-718-1

Query Match 99.5%; Score 6939; DB 8; Length 6975;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6972; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

Qy 1 ccattggtatggcaggtgtctgcctgcctgctatgagcagatgacacccatttgcc 60
Db 1 CCATGGATATGGCAGGTGTCTGCCCTGCTATGATGGCGATGACACCCATTGGCCC 60
Qy 61 catatctgacgatttgacatgtgatatttaacatgtgacatgatttaacattgttt 120
Db 61 CATATCTGTACGATTTGACATGTGATATGATTTAAACATGTGACATGATTTAACTTTT 120
Qy 121 aatactgttgccatcattaccataatttagtaacgcatttagtaacgcatttgtaaaaaat 180
Db 121 AATACTGTGCCATCATTAACCAATAATTTAGTAACGCATTTAGTAACGCATTTGTAAAAAT 180
Qy 181 cattgcgccctttatgtatcatcatgaatagaattatttatgtattctgattattgt 240
Db 181 CATTCGCCCCCTTTATGTATCATATATGAATAATATATGATGATGATGATGATGATGAT 240
Qy 241 atcagaatgggtgatctatgatgatgcctacagagttgatttgggttaactcactctatg 300
Db 241 ATCAGAATGGTGATGCTATATGATGATGCTACGAGTTGATTTGGTTAATCACTCTATG 300
Qy 301 atttgatatattttgaaactaatctatttgacttaataatcaccatattggttataatttagca 360
Db 301 atttgatatattttgaaactaatctatttgacttaataatcaccatattggttataatttagca 360

Db 301 ATTGATATATTTTGAACCTAATCTATTGACTTAATACACATATGTTATTAATTTAGCA 360
 Qy 361 taatggttaggtcttttgaataatacaatcacaataattgttctactgttactaccatgct 420
 Db 361 TAATGTTAGGCTTTTGTAAATAATCACATCGCAATATTGTTCTACTGTTACTACCATGCT 420
 Qy 421 tgaatgacgatcccaatcacagatcttcaagtgatgtgtttgtatatacagcacattta 480
 Db 421 TGAATGACGATCCCAATCACAGATTCATTCAAGTGATGTGTGTATACGACCATTTA 480
 Qy 481 cccataatttcaatcaaatccatgctatgctacgatgtatcatcttttttaaggttaaacac 540
 Db 481 CCCTAATATTCAATCAATCCATGCTATGTCAGCATGTATCATTTTTTTTAAGGTAAACAC 540
 Qy 541 catgaatcacatctataaagtcatcttttaacaaagccacagggcacatttatggcagtgcc 600
 Db 541 CATGAATCATCTATAAGTCATCTTTTAACAAAGCCACAGGCACATTTATGGCAGTGGC 600
 Qy 601 agagtacggcaaatcccaacagcaac-gggggggggtagctgtgctacagggcaagttggca 659
 Db 601 AGAGTACGCCAAATCCACAGCACGGGGGGGGTAGCTGTGTACAGGGCAAGTTGGCA 660
 Qy 660 gtgatcaccttgagctttgcccgtatgcccgcgtcgctgtctcgtgtgacgctgacg 719
 Db 661 GTGTATGCACTCTGAGCTTTGCCCGTATTGCCGCGCTCGCTGTCTCTGATCGGTGCAA 720
 Qy 720 cgtcagtgagctgttctcctcaaaaaaagatatcccaaaacatatcgcaattggtgaac 779
 Db 721 CGCTACGTGCGAGTGTATGCTCMAAAAAGATACCAACATATCGCAATTTGGTGAAC 780
 Qy 780 aaaaacggcaagacgctgagcaactcccaagcgagcggtgacggtgacgcatgctattg 839
 Db 781 AAAAAACGCAAGACGCTCAGGCACTGCCAAGCGGACGGGTGATCGAGCCATTGCTATTG 840
 Qy 840 gtgaaaaagctaaacgacagggcggttaagccatcgccatcggttagtagtaataaaactg 899
 Db 841 GTGAAATGCTAACGCACAGCGCGGTCAAGCCATCGCCATCGTAGTAGTAATAAAACTG 900
 Qy 900 tcaatggaagcagtttggaataagtaagtaaccagctacggtgcaagagtcacatgcgca 959
 Db 901 TCATGGAAGCAGTTTGGATGAAGATAGTACCGATGCTACGGGTCAAGAGTCCATGCCA 960
 Qy 960 tcggtggtgataaaggctagtgctcgtcgtatgcccattcggttagtgagcttac 1019
 Db 961 TCGGTTGGTATGATAAAGCTAGTGTGTGCTGCTGATGCTGCTGCTGCTGCTGCTGCT 1020
 Qy 1020 atttcttgatcagcatgtaattcctaaacatccgaaggtactctgtattaaacgactta 1079
 Db 1021 ATTTGCTTGATCAGCATGTTATCTTAACATCCGAAGGTACTCTGATTAACGATCTTA 1080
 Qy 1080 ttaacggccatcagtaataaagaaaatacgaagctcaaaagataatgatgataaaata 1139
 Db 1081 TTAACGGCCATCGATATTAAAGAAATACGAAGCTCAAGGATAATGATTAATAATATA 1140
 Qy 1140 gaagcaaacccgcaagcggaacccagctactcagtgagggagccatgcatatgcacagg 1199
 Db 1141 GAGCGCAACCCGACAGCGGACCGGACGACTGCAAGTGGGACCATGCTATATGCACAGG 1200
 Qy 1200 gtcattttlccacagcctttgttacacggggcaacagctaaagtgccctattccttgagag 1259
 Db 1201 GTCATTTTCCACAGCCTTTTGTACAGCGGCAACAGCTAAAAGTGCCTATTCTTTGGCAG 1260
 Qy 1260 tgggttccgcccacagccgagggccaatctacaatcgctattggttcttgatgcaaat 1319
 Db 1261 TGGGTCTTGGCGCCACAGCGGAGGGCAATCTACAATCGCTATTGTTCTGTATGCAACAT 1320
 Qy 1320 ctgctgtttgggagcagtagcccttgggtgcaggttactcgtcagctacagggcagta 1379
 Db 1321 CTAGCTGTTGGGCGATAGCCCTTTGGTGCAGGTACTCGTGTCTGCTGCTGCTGCTGCT 1380
 Qy 1380 ttgcccctaggtcaaggttctgttgcactcagagtgataataattctagaccgacctata 1439
 Db 1381 TTGCCCTAGGTCAAGGTTCTGTGTCTCAGAGTGATTAATAATTTCTAGACCGGCTATA 1440

Qy 1440 caccataataccagcactagaccacaaagtttcaagccaccataataacgaagggggtc 1499
 Db 1441 CACCAATACCCAGGACCTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGGGGGTTC 1500
 Qy 1500 cactttccattggttagtaactctatacaacgtaaaatcatcaatgtcgggtcaggtgta 1559
 Db 1501 CACTTTCCATTGTTAGTAACTCTATCAAAACGTAATAATCATCAATGTCGGTGCAGGTGTTA 1560
 Qy 1560 ataaaacccgatcggttcaatgtggcacagctagaagcgggtgggaagtggggtaaggggc 1619
 Db 1561 ATAAAACCGATCGGTCATATGTGCCACAGCTAGAACGGTGGTGAAGTGGCTTAAGGAGC 1620
 Qy 1620 gtagaattacttttccagggtgataaactactgtacgtataaaataggttttggataata 1679
 Db 1621 GTAGATTACTTTTCAGGGTGATGAATAACAGTACTGACGTAAATAAGTTTGGATTAATA 1680
 Qy 1680 ctttaactataaaggtggtgcagagacccaacgataaaccgataataatcgggtgtg 1739
 Db 1681 CTTTAACTATTAAAGGTGGTGCAGAGACCAACCGATTAAACCGATTAATAATATCGGTGTGG 1740
 Qy 1740 taaagagggctgaataatgtgctgaaagttaaactgtctaaacttttaacaaacttta 1799
 Db 1741 TAAAGAGGCTGATTAATAGTGGTCTGAAAGTTAACTTGTCTAAACCTTTAAACAACTTTA 1800
 Qy 1800 ctgaggtgataaactacatttaataatgcacaacacagtttaaggttaggtagtagta 1859
 Db 1801 CTGAGGTGAATTACAACTACATTAATAATGCCACACACGTTTAAAGTAGTAGTAGTA 1860
 Qy 1860 gtactcagctgaattattgagtgatagtttaacctttaccagcccaatacagggcgctc 1919
 Db 1861 GTACTACAGCTGAATATTAGTGTGATAGTTTAACTTTTACCAGCCCAATACAGGCGATC 1920
 Qy 1920 aaagcacaagcaaaacccgtctatggcgttaattggcgttaataatggcgttttgcgagatg 1979
 Db 1921 AAAGCACAGCAAAACCCGTCTATGGCGTTAATGGGGTGAAATTTGCTTTCGTGAGATG 1980
 Qy 1980 caacagcgcaatcgccactactctgtattaccagagataaaatctggccttctcgcagatg 2039
 Db 1981 CANCAGAGCAATCGGACACTACTCTGATTTACCAGAGATAAAATTTGGCTTTGCTGAGATG 2040
 Qy 2040 gtgattgtgataaagcaaacagcaccatatttggataaaacacacttaagtgaggtgag 2099
 Db 2041 GTGATGTTGATGAAACAAAGCACCATTATTGGATTAATAAACAACTTAAAGTGGGTAGT 2100
 Qy 2100 ttgcaattaccatagacaaatggcattgtgcagggttaaataaaagatcgagtaactctgca 2159
 Db 2101 TTGCAATTTACCATAGACATGCAATGCAATGATGCAAGTAAATAAAAGATCAATATCTTGCCA 2160
 Qy 2160 aaggtagcagtgctaacgactgaggttaccatcgaaacagctcaagcccaagcctactt 2219
 Db 2161 AAGGTAGCAGTGTCTAACGATGCGGTTTACCATCGACAGCTCAAAAGCCCAAGCCCTACTT 2220
 Qy 2220 taacagcagcgctggtcagctcagtgctcacacccactgaaatacagttgagtgaagagtg 2279
 Db 2221 TAAACCGAGCGCTGCGATCAGTGTACACCTACTCAATATATCAGTTGATGTAAGAGTG 2280
 Qy 2280 gcaatgttaccgcccccaacttacaacatttggcgtgaaaccccaagcttaacagtgatg 2339
 Db 2281 GCAATGTTTACCGCCCCAACCTTACAACATTTGGCGTGAACCCACCGAGCTTAAACAGTATG 2340
 Qy 2340 gcaatgttaccgcccccaacttacaacatttggcgtgaaaccccaagcttaacagtgatg 2399
 Db 2341 GCATAGTGATTAATTTTGTGTTAAGGTAGTGTGAGCAACAAATAGCTTAGTTACGCGCG 2400
 Qy 2400 aacattggcgaagctatctaaatgaagtcatacgaacggctgacagtgctctacaagct 2459
 Db 2401 AACATTTGGCAAGCTATCTAAATGAAGTCAATCGAAGCGCTGACAGTGTCTTCAAAAGCT 2460
 Qy 2460 ttacgggttaagaagaagacgatgatgacccaacgctatcacccgtggcttaagatacga 2519
 Db 2461 TTACCGTTAAAGAAAGACGATGATGACGCCAACGCTATCACCGTGCCTTAAAGATACGA 2520

Db 4681 AAGGGCAGGCAAGCAACAACTACAGAGGCTATGTGGATGCTGATGGCAATTAAGGTCA 4740
 Qy 4740 tctatgacagtaccgataaacaagtactatcaagccaaataatgatggcacagtgtataaaa 4799
 Db 4741 TCTATGACAGTACCGATACAAAGTACTATCAAGCCAAATAATGATGGCACAGTTCGATAAAA 4800
 Qy 4800 ccaagaagtgtccaaagacaaaacttggctggcccaagcccaaacccccagatggacattgg 4859
 Db 4801 CCAAGAAGTGTCCAAAGACAACTGGTGGCCAAAGCCCAAGCCCAAGATGGCATTTGG 4860
 Qy 4860 ctcaaatgaatcctcaaatcagtcattcaacaagaacaagttaaatgatcctcaataaaaagc 4919
 Db 4861 CTCAAATGAATGTCAAAATCGTTCATTAACAAACACAAAGTAAATGATGCCAATAAAAAGC 4920
 Qy 4920 aaggcctaatgaagacaaacgccttctgttaaaggacttgaagaaagccgccttcttgataaca 4979
 Db 4921 AAGGCATCAATGAAGCAACAGCGCTTTGTAAAGGACTTTGAAAGAGCCGCTTCTGTATAACA 4980
 Qy 4980 aaacccaaacccgcgcagtaactgtgggtgatttaaatgcgcttggccaaacacccgcctga 5039
 Db 4981 AAACCAAAACCGCGCAGTAACTGTGGGTGATTTAAATGCCGTTGCCCAAAACACCGCTGA 5040
 Qy 5040 cctttgcaggggtacagcgcaaacgcgtctaaacaaactggcgagactttgacactcaaaag 5099
 Db 5041 CCTTGCAGGGGATACAGGCACAAAGCGCTTAAAAACTGGCGGAGACTTTGACATCAAAAG 5100
 Qy 5100 gtgggcaaacagacacataagctaacgcgataataaacatcggtgtggtagcagtgactg 5159
 Db 5101 GTGGGCAACAGACACCAATAAGCTAACCGATAAATACATCGGTGTGGTAGCAGGTACTG 5160
 Qy 5160 atggcttcactgtcaaaactgtccaaagacctaaccacttcttaacagcgtttaatgcaggttg 5219
 Db 5161 ATGGCTTCATCTCAAACTTGGCCAAAGACCTAACCAATCTTAACAGCGTTAATGACAGGTG 5220
 Qy 5220 gcacccaaattgatgacaaaggcgtgtctttgtgactcaagcgttaagccaaagcaa 5279
 Db 5221 GCACCAAAATGTATGACAAAGGCGGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAA 5280
 Qy 5280 acacccctgtcctaagtcgcaatggcgtgacgtggtggcaaggctcaatgtaagtgtg 5339
 Db 5281 ACACCCCTGTGCTAAGTGCCTAATGGGCTGCACCTGGGTGGCAAGGTCATCAGTAATGTGG 5340
 Qy 5340 gcaaggcacaataagatacoga-cgctgccaatgtacaaacagtaaacgaagtacgcaac 5398
 Db 5341 GCAAGGACACAAAAGATACCGACCGCTGCCAATGTACACAGCTTAACAGAAATACGCAAC 5400
 Qy 5399 ttgttggtctgtgaatgctgtgtaataagctgaacgctgaacgctcaatcaggttaaacattg 5458
 Db 5401 TTGTTGGGTCTTGGTAATGCTGTGTAATGTAATACGCTGACGGCAATCAGGTAAACATTGCC 5460
 Qy 5459 gacatcaaaaagaccacaaattccaggttcatctaatccgcgactgtcatcaagaagcgc 5518
 Db 5461 GACATCAAAAAGACCCAAATTCAGGTTCATCATCTAACCGCACTGTCAATCAAGCAGGC 5520
 Qy 5519 acggtacttggcgttaaaagtaataacgataaccgaaacacttgcactgtgtggtatacaa 5578
 Db 5521 ACGGTACTTGGCGTAAAGGTAATACGATACCCGAAACACTTGCACATGTGTTATACAA 5580
 Qy 5579 gtggcgctggataaagacgcaacgctaacgctgaatgaagcaatgtttgggtcaaaacc 5638
 Db 5581 GTGGCGGTGGATTAAGACGCAACGCTAACGCGGATTAAGCAATGTTGGGTCAAAAACC 5640
 Qy 5639 caaaaagatggcagcaaaaagccctgtcgcacttataacgcgcaggtcagaccacac 5698
 Db 5641 CAAAAGATGGCAGCAAAAAGCCCTGTCCGCACTTATAACGCGCGCAGGTACACCAAC 5700
 Qy 5699 tattgaccaacaacccgcagaaagccattgacagaaataatgaacaagggtatccgcttc 5758
 Db 5701 TATTGTACCACAAACCCCGCAGAACCCATTGACAGAAATAATGAACAAAGGTATCCGCTTC 5760
 Qy 5759 ttccatgtcaacgataggcaataagagcgtgtgtacaaggcgtaacagcgattgactca 5818
 Db 5761 TTCCATGTCAAGGATGGCAANTCAAGAGCCTGTGGTACAAAGGCGGTAAACGGCATGACTCA 5820

Qy 5819 atgacctcaggcaagcactcagtggcqataggtttccaggcccaaggcagatggtgaagcc 5878
 Db 5821 AGTGCTTCAGCCAAAGCACTCAGTGGCGATAGGTTCCAGCCAAAGCAGATGTTGAAGCC 5880
 Qy 5879 gccgttgccatagcgagacaaacccaagcaggcaacccaatccatccatcgccatcggttgataac 5938
 Db 5881 GCCGTTGCCATAGGCAGACAAACCAAGCAGGCAACCAATCCCATCGCATCGGTGTATAAC 5940
 Qy 5939 gcaaaagccagggcgatcaatccatcgccatcggttacaggcaatggttagcaggtgaag 5998
 Db 5941 GCACAAAGCCAGGCGGATCAATTCATCGCCATCGGTACAGGCAATGTTGGTAGCAGTAAG 6000
 Qy 5999 cactctgtgtccatcgcgacccaagcactgtttaaggctgataaacaggttacagtgtggtt 6058
 Db 6001 CACTCTGTGCTGCCATCGCGACCCCAAGCACTGTTAAAGGCTGATAACAGTTACAGTGTGGT 6060
 Qy 6059 aataacaacagtttacggatgacacccaacccagatgtcttgggtgggcaataaacatc 6118
 Db 6061 AATAACAACAGTTTACCGATGCCACTCAAAACCGATGCTTTGGTGTGGCAATTAACATC 6120
 Qy 6119 accgtgacccgaagttaactcggttgccttaggttcaaaactctgccaatcagtcaggcaca 6178
 Db 6121 ACCGTGACGAAAGTAACCTGGTTGCCCTTAGGTTCAAACTCTGCCATCAGTCAGGACACA 6180
 Qy 6179 cacgcaggcacacaagccaaataatctgacgcagcagcaggtacacacacacacagcaggt 6238
 Db 6181 CACGCAGCACACAAAGCCAAATAATCTGACGGCACAGGATCAACACCAACACACAGCAGGT 6240
 Qy 6239 gcaaccgtacggttaaaaggcttgcctggacaaacgcggttggctgcggtctcgcgtgggt 6298
 Db 6241 GCAACCGGTAGCGTTAAAGGCTTGTGACACAAACGCGGTTGGTGGGTCTCGCTGGGT 6300
 Qy 6299 gctcaggtgtgcaacgcgctatccaaatgtgcagcaggtgaggtcagtcagtcagccagcgc 6358
 Db 6301 GCCTCAGGTGTGAACGCGCTATCCAAATGTGCAGCAGGTGAGGTGAGTCCACCCACAGC 6360
 Qy 6359 accgatcggtcaatggtagccagttgtacaaagccaccccaaaagcattgccaacgcgaacc 6418
 Db 6361 ACCGATGCGGTCAATGTGTAGCCAGTTGTACAAGCCACCCCAAGCAATGTCACCAAGCAACC 6420
 Qy 6419 aatgagcttgaccatcgatatccacccaaacaaataaggccaaatgcaggatatttcataca 6478
 Db 6421 AATGAGCTTGACATCGTATCCACCAAAACGAAATAAGGCCAATGACAGGATTTTCATCA 6480
 Qy 6479 gcatggcgatggcctcatgcacaagcctacattcctgagcagatccatggttaccggg 6538
 Db 6481 GCGATGGCGATGGCGTCCATGCCACAAAGCCTACATTCCTGGCAGATCCATGTTACCAGG 6540
 Qy 6539 ggtattgcccacccacaaacgggtcaaggtgcgggtggcagtgggactgtcgaagcgtcgat 6598
 Db 6541 GGTATTGCCACCCACAAACGCTCAAGGTGGGTGGCAGTGGGACTGTCCGAAGCTGCGGAT 6600
 Qy 6599 aatggccaatgggtatttaaaatcaatgttcagccgataccacagggccatgtaggggcg 6658
 Db 6601 AATGGTCAATGGGTATTTAAATCAATGTTTACCGCATACCCCAAGGCCATGTAGGGGCG 6660
 Qy 6659 gcatgtgtgaggttttccacttttaagccataaaatgcgaagatttttacttaaaatcaa 6718
 Db 6661 GCAGTTGGTGCAGGTTTTCATTTTAAGCCATAAATGCAAGATTTTACTTAAATCAAA 6720
 Qy 6719 tctcacatagttgtat-aaacagcatcagcatcagtcataattactgactgactgactt 6777
 Db 6721 TCTCACCATAAGTTGATAAAACAGCATCAGCATAGTCATATTTACTGATGCTGATGTTT 6780
 Qy 6778 ttatcacttaaacacttttaccgctcaagtgattctcttccaccatgaccacaaatgcga 6837
 Db 6781 TTTATCACTTTAAACATTTTACCGCTCAAGTGATTTCTTTTACCATTGACCAANTCGCCA 6840
 Qy 6838 ttgatcataggttaaacctttaggttaaaattttatcaaatgttagttgtagtatggttaaa 6897
 Db 6841 TTGATCATAGTAACCTTATTGATTAATTTTATCAATGTAGTTGTTAGATATGGTTAA 6900

QY 6898 attgtgcatgaccacaaatgaccgatttatcccgaaaaatttctgattgatccggtt 6957
|||||
Db 6901 ATTGTGCCATTACCAAAAAATGACCGATTATCCCGAAAAATTTCTGATTATGACCGTT 6960
QY 6958 gacctgcaggtcgac 6972
|||||
Db 6961 GACCTGCGAGTGCAC 6975

RESULT 11

US-08-431-718A-1
; Sequence 1, Application US/08431718A
; GENERAL INFORMATION:
; APPLICANT: Sasaki, Ken
; APPLICANT: Harkness, Robin E
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431.718A
; FILING DATE: 01-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-429
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-431-718A-1

Query Match 99.5%; Score 6939; DB 8; Length 6975;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6972; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 1 ccatgatatggcgagggtgtgtcgtccgtccgtatgatggcgatgacacccattggccc 60
|||||
Db 1 CCATGGATATGGCAGGTGTGTCGCGCTGCCGTATGATGGCGATGACACCCCATTTGCC 60
QY 61 catatctgtacagatttgacatgtgatattgatttaacatgtgacatgatttaacattgttt 120
|||||
Db 61 CATATCTGTACGATTGTGACATGTGATGATTAAACATGTGACATGATTAAACATTGTTT 120
QY 121 aatactgttgccatcattaccataatttagtaacgcatcttagtaacgcatcttgtaaaaaa 180
|||||
Db 121 AATACTGTGGCCATCATTAACCAATATTAGTAACGCATTTAGTAACGCATTTGTAATAAT 180
QY 181 cattgccccttattgtgtatcatatagaataattatgattgtatctgattattgt 240
|||||
Db 181 CATTGCGCCCTTATTGTGTATCAATGAATATATGATTGTTATCTGATTATTCT 240
QY 241 atcagaatggtgatcgtatatgatgatgcctacgaggttgatttgggttaataacactctatg 300
|||||

Db 241 ATCAGAAATGGTGATGCTATATATGATGATGCTACGAGTTGATTTGGGTTAATCACCTCTATG 300
QY 301 atttgatatatttgaacactaatctattgacttaaatcaccataatggttataattagca 360
|||||
Db 301 ATTTTCATATATTTTCAAACTAATCTATTTGACTTAAATCACCATATGGTTTATATTTAGCA 360
QY 361 taatggttaggctttttgtaaaaaatcacatcgcaatattgttctactgttactaccatgt 420
|||||
Db 361 TAATGCTAGGCTTTTGTAAAAATCACATCGCAATATTGTTCTACTGTTACTACCATGCT 420
QY 421 tgaatgacgacatcccaatcacagattcattcaagtgtatgtgttttatcagcacattta 480
|||||
Db 421 TGAATGACGATCCCAATCACCAGATTCAATCAAGTGATGTGTTGTATACGCCACCATTTA 480
QY 481 cccataattttcaatcaaatgccttatgtcagcatgtatcatttttttaaggttaaacccac 540
|||||
Db 481 CCTAATTTTCAATCAAAATGCCATGTATGTCACGATGTATCATTTTAAAGGTAAACCCAC 540
QY 541 catgaatcacatctataaagtcatcttttaacaaagccacagcacatttatggcagtggc 600
|||||
Db 541 CATGAATCACATCTATAAGTCTATCTTTAACAAAGCCACAGGCACATTTATGCGAGTGC 600
QY 601 agagtacgccaatccacagcac-gggggggggtagctgtctacagggcgaagtggca 659
|||||
Db 601 AGAGTACGCCAAATCCACAGACGGGGGGGGTAGCTGTCTACAGGGCAAGTTGGCA 660
QY 660 gtgtatgcaactctgagctttgcccgtattgcgcgcgtcgtcgtcctcgtgacggtgca 719
|||||
Db 661 GTGTATGCATCTGAGCTTTGCCCGTATTGGCCGCTCGCTCTCTCTGTCGTATCGGTGCA 720
QY 720 cgtcagtggtgagtgcttatgtctcaaaaaaagataccaaacatctcgaattggtgaac 779
|||||
Db 721 CGCTCAGTGGCAGTGTATGCTCAAAAAAAGATACCAACATATCGCAATTTGGTGAAC 780
QY 780 aaaaacgacgaagcgtcagcactgcgaagcgagcgtgatcgagcattgctattg 839
|||||
Db 781 AAAACGACCCAAAGCGCTCAGGCATGCGCAAGCGGACGGTGATCGAGCATTTGCTATTG 840
QY 840 gtgaaaatgtcaacgcacagcggtgtcaagccatcgcaatcggtgtagtagtaataaac 899
|||||
Db 841 GTGAAATGCTAACGCACAGGGCGGTCAAGCCATCGCCATCGGTAGTAGTAATAAATCTG 900
QY 900 tcaatggaagcagtttggataagataggtaccgatgtctacgggttcaagagtcacgcga 959
|||||
Db 901 TCAATGGAAGCAGTTTGGATAAGATAGTAGTACCGGTCTACGGGTCAAGAGTCCATCGCCA 960
QY 960 tcggtggtgatgtaaaaggctagtggtgatgctcgtatgcccacgtcggttagtgatgacttac 1019
|||||
Db 961 TCGGTGGTGATGTAAGGCTAGTGGTGATGCTCGCTCGCATTCGCGTGTGATGACTTAC 1020
QY 1020 atttcttgatcagcatggttaactcctaaacatccgaaggtactctgtatgaacgatctta 1079
|||||
Db 1021 ATTTCTTGATCAGCATGCTAATCTTAACATCCGNAAGGTACTCTGATTACGATCTTA 1080
QY 1080 ttaacggccatgcagttattaaagaataacgaagctcaagagataatgatgtaaaata 1139
|||||
Db 1081 TTAACGGCCATGCAGTATTAAAAAGAAATACGAAGCTCAAAAGGATAATGATGTAAAAATA 1140
QY 1140 gacgcacaacccgaagcggacacgcccagtaactgagtgagtgagggccatgtcatatgcaagg 1199
|||||
Db 1141 GACGCACAACCCGAAGCGGACACGCGAGTACTGCAAGTGGGAGCCATGTATATGACACAG 1200
QY 1200 gtcaattttccaaacgctttggtacacggtgcaacagctaaaagtgctcttctctggcag 1259
|||||
Db 1201 GTCAATTTTCCAAACGCCCTTTGGTACACGGGCAACAGCTAAAAGTGCCTATTCTCTGGCAG 1260
QY 1260 tgggttcccgccacagccgaggggccaatctacaatcgctattggttctgatgcaaac 1319
|||||
Db 1261 TGGGTCTTCCGCCACACAGCCGAGGGCCAACTTACAATCGCTATTGGTTCTGTATGCAACAT 1320
QY 1320 ctagctcgttggagcgatagcccttggtgagggtagctcgtcgaagtcacagggcagta 1379
|||||
Db 1321 CTAGTCGTTGGAGCGATAGCCCTTGGTGCAGGTACTCGTCTCAGTACAGGGCAGTA 1380

QY 1380 ttgccctaggtcaaggtctctgttgcactcagcagtgataataattctagacggcctata 1439
 Db 1381 TTGCCCTTAGGTCAAGGTTCTGTGTACATCAGAGTGATATAATTTAGACGGGCTATA 1440
 QY 1440 caccataatccaggcacttagaccccccaagtttcaagccacccaataatcgaagcggttc 1499
 Db 1441 CACCAATAATACCCAGGCACTAGACCCCAAGTTTCAAGCCACCAATATACGAAGCGGGTC 1500
 QY 1500 cactttccattgttagtaactctatcaaacgttaaaatcatcaatgtcgggtcagtggtta 1559
 Db 1501 CACTTTCATTTGTAGTAACTCTCATCAACGTAAATATCATATGTGCGGTGAGGTGA 1560
 QY 1560 ataaaccgctgcggttcaatgtggtgcacagctagaagcggtgtgaagtgggtgaaggaac 1619
 Db 1561 ATAAACCGATGCGGTCAATGTGTGCACAGCTAGAACGGTGTGTAAGTGGGCTAAGGAGC 1620
 QY 1620 gtagaattacttttcagggtgtagataaacagtagcagctaaaggttgaaggttttggaataa 1679
 Db 1621 GTAGAAATTTACTTTTCAGGGTGTATGAACAGTACTGACGTAAANATAGGTTGGGATAA 1680
 QY 1680 ctttaactattaaaggtgtgtgcagaccacacgcattaaaccgataataatcgtgtgtg 1739
 Db 1681 CTTTAACTATTAAAGGTGTGTGCAGAGACCAACGCATTAACCGATAATAATATCGGTGTG 1740
 QY 1740 taaagaggctgataatagtggttgtaaaagttaaaacttgctaaacttttaacaaacttta 1799
 Db 1741 TAAAGAGGGCTGAATAATAGTGGTCTGAAGTTAAACTTGTCTAAACTTTAAACAACTTA 1800
 QY 1800 ctgagggtgaatacaactacataataatgccacaaccacagtttaagggtaggttagtagta 1859
 Db 1801 CTGAGGTGAATACAACTACATTAATTAATGCCACACACAGTTAAAGGTAGGTAGTAGTA 1860
 QY 1860 gtactaacgctgaattatttgagtgatagtttaaaactttaccagcccaatacagcgagtc 1919
 Db 1861 GTACTACAGCTGAATTAATGTAGTGATAGTTAACTTTAACCTTTACCAGCCCAATACAGCGAGTC 1920
 QY 1920 aaagcacaaagcaaacccctctatggcttaagtgggtgaaggtttactaataatgcagaaa 1979
 Db 1921 AAAGCACAAAGCAAAACCGTCTATGGCTTAATGGGTGAAGTTTACTTAATGCAGAAA 1980
 QY 1980 caacagcagcaactcggcactactcgtattaccagagataaaattgcttgcctcgaagt 2039
 Db 1981 CAACAGCAGCAATCGGCCTACTCTGATTTACAGAGATAAAATTTGGCTTGTCTCGAAGTG 2040
 QY 2040 gtgattgtgataaaacaagaacacacataatttggataaaaaaaccttaaaagtgggtagt 2099
 Db 2041 GTGATGTTGTAAGAAACACACACCATATTTGGATAAAAAACAACTTAAAGTGGGTAGTG 2100
 QY 2100 ttgcaattaccatagacaatggcattgatgcaggtataaaaaagatcagtaacttgcca 2159
 Db 2101 TTGCAATTAACCATAGACATAGCAATGGCAATGTATGAGGTAAATAAAGATCATGTAATCTTGCCA 2160
 QY 2160 aaggtagcagtgactaacgatcggttaccatcgaacagctcaaaagccgcgaagcctactt 2219
 Db 2161 AAGGTAGCAGTGCTACGATGCGGTACCATCGAACAGCTCAAAAGCCGCAAGCCTACTT 2220
 QY 2220 taacagcagcgtggtgcatcagtgatcaacotactgaaataatcagttgatgctaaagtg 2279
 Db 2221 TAAACGCAAGGCTGGCATCAGTGTGCACACCTACTGTGAATATATCATGTTATGCTAAGAGTG 2280
 QY 2280 gcaatgtaccgcccccaacttaaacattggcgtgaaacaccagcgttaacagtgatg 2339
 Db 2281 GCAATGTTACCGCCCCCAACTTCAACATTTGAGGTAGTGGTACGAACAATAAGCTTAGTACCGCG 2340
 QY 2340 gcaactagtgataaatttagtttaagggtagtggtacggaacaataagcttagttaccgcg 2399
 Db 2341 GCCTAGTGATAAATTTAGTGTTAAGGGTAGTGGTACGAACAATAAGCTTAGTACCGCG 2400
 QY 2400 aacatttggcagcgtctataatgaagtcaatcgaacgctcagcgtgctctacaagct 2459
 Db 2401 AACATTTGGCAAGCTATCTTAATTAATGAAGTCAATCGAACGGCTGACAGTGTCTACAAAGCT 2460

QY 2460 ttaccgttaagaagaagacgatgatgacgcccaacgcgtatcacgcgttggttaagaatacga 2519
 Db 2461 TTACCGTTAAGRAAGAAGACGATGATGACGCCAAGCGTATCACCGTGGCTAAAGATACGA 2520
 QY 2520 caaaaaatgcggcgagtcagcatcttaaaactcaaaagtaaaaaacggtcttaacggttg 2579
 Db 2521 CAAAAAATGCGCGCGCAGTCAGCATCTTAAACCTCAAAAGGTAAACCGGTCTTACGGTTG 2580
 QY 2580 ctacaaaaaagatggttacggttacctttgggtcttagccaagatagcgggtctctgacattg 2639
 Db 2581 CTACAAAAAAGACATGTTACGGTTACCTTTGGGCTTTAGCCCAAGATAGCGGTCTGACCAATTG 2640
 QY 2640 gcaaaagcaccctaaacaacgatggttgacgtgtaaaagataccaaagaacaataccaag 2699
 Db 2641 GCAAAAGCACCTAAACAACGATGGCTTGACTGTTAAAGATACCAACGAACAATAATCCAAG 2700
 QY 2700 tgggtgtaattggcattaaatttactaatgtgaatggttagtaatccagggtactggtcattg 2759
 Db 2701 TCGGTGCTAATGGGCATTTAATTTTACTAATGTGAATGTTAGTAAATCCAGGTACTGSCATTG 2760
 QY 2760 caaataccgctgcgcatcaccagagataaaattggctttgtggtctctgagtggtgcagttg 2819
 Db 2761 CAATACCGCTCGCATTTACCAGAGATAAAATTTGGCTTTGCTGGTCTGATGGTGCAGTTG 2820
 QY 2820 atacaaaaaacaccttaacttctgatcaagaacagctacaagttggtcaatggttaagattacca 2879
 Db 2821 ATACAAACAAACCTTTATCTTGATCAAGACAAAGCTACAAAGTTGGCAATGTTAAGATTACCA 2880
 QY 2880 acactggcattaaqcgaggtggttaagccatcacagggtctcccccaacactgocctagca 2939
 Db 2881 ACACCTGGCATTAACGACGCTGTTAAAGCCATCACAGGGCTGTGCCCAACACTGCGCTAGCA 2940
 QY 2940 ttgcgagataaaagtgccgcaacacatagaaactgggcaatatacaatccaagacaagaagcaaat 2999
 Db 2941 TTGCCGATCAAAAGTAGCCGCAACATAGAACTGGGCAATACAAATCCCAAGACAACAAAT 3000
 QY 3000 ccaagctgccagcattaatgatataataacaggctttaacccttaaaaaataataaca 3059
 Db 3001 CCNACGCTCCAGCATTAATGATATATTAATTAACAGGCTTTAACCTTAAAAATAATAACA 3060
 QY 3060 acccaattgacttctcactctatgacatttggtaacttggcaatggcaatgccacca 3119
 Db 3061 ACCCAATTGACTTTGTCTCCACTTATGACATTTGTGACTTTGCCAATGCAATGCCACCA 3120
 QY 3120 cggccacagtaacccctgataccgctcaacaaaccagtaaaagtggtatataatgattgaaatg 3179
 Db 3121 CCGCCACAGTAACCCATGATACCGCTAACAAACCAAGTAAGTGGTATATGATGTGAATG 3180
 QY 3180 tggatgatacaacacattcattcaacaggcactgatacaataaaaaaaccttggtgcacaaa 3239
 Db 3181 TGGATGATACACCAATTCATCTAACAGGCACATGATCACAAATAAAAAAATTTGGCGTCAAAA 3240
 QY 3240 ccaacaaactgaacaaaaaagtgcttaattggttaatacagcaacttaacttaacttaact 3299
 Db 3241 CCACCAAACTGAACAAACAAAGTGTAAATGTTAATACGCACTAATCTTAATGTTAACT 3300
 QY 3300 ctagtgtgaagtgcctctttaaagcgaagacatcgccgaaaatctaaacacccctag 3359
 Db 3301 CTAGTGATGAAGATGCCCTTTGTTAACGCCAAAGACATTCGCCGAAAAATCTTAACACCCCTAG 3360
 QY 3360 ccaaggaaattcacacacacacaaaggcagcagacacccgctctacaaaccttaaccgtta 3419
 Db 3361 CCAAGGAAATTCACACCAACCAAGGACAGACACACCGCCCTACAAACCTTTACCGTTA 3420
 QY 3420 aaaaggtagatgaataataatgctgatacgcacaaacgccaatacgcgtgggttcaaaaga 3479
 Db 3421 AAAAGGTAGATGAATAATAATGCTGATAGCCCAACGCAATCATCCGCTGGGTGCAAAAGA 3480
 QY 3480 acgcaataatacagtgtaacacacccctaaactcaaaagtgaaacggtcttaataataaaa 3539
 Db 3481 ACGCAAAATAATCAAGTCAACACCCCTAACACTCAAAAGGTGAAACCGGCTTAAATATTTAAAA 3540
 QY 3540 ccgacaaaaatgggtacggttacctttggtcattacacacacacaaagcggtctttaaagccggca 3599

Qy	5759	ttcctgtcaacgattgccaatacaagagcctgtgtgtacaaggcgctaaacggcattgactca	5818
Db	5761	ttccattgtcaacgattgccaatacaagagcctgtgtgtacaaggcgctaaacggcattgactca	5820
Qy	5819	agtcctcaggccaagcactcagttggcgatagggtttccaggccaagcgagatggtgaagcc	5878
Db	5821	agtcctcaggccaagcactcagttggcgatagggtttccaggccaagcgagatggtgaagcc	5880
Qy	5879	gccgttgccataggcagacaacccaagcaggcaaccaatccatcgccatcggtgataac	5938
Db	5881	gccgtttgccataggcagacaacccaagcaggcaaccaatccatcgccatcggtgataac	5940
Qy	5939	gcacaagccacggggcgatccaatccatcgccatcggtcagggcaatgtggtagcaggaag	5998
Db	5941	gcacaagccacggggcgatccaatccatcgccatcggtcagggcaatgtggtagcaggaag	6000
Qy	5999	cactctgggtgccatcggcgagcccaagcaactgtaagctgtatacaggttacagttgggt	6058
Db	6001	cactctgggtgccatcggcgagcccaagcaactgtaagctgtatacaggttacagttgggt	6060
Qy	6059	aatacaaacccagtttaccgatgccactcaaacccgatgcttcttggttggggcaataacatc	6118
Db	6061	aatacaaacccagtttaccgatgccactcaaacccgatgcttcttggttggggcaataacatc	6120
Qy	6119	accgtgacgaagaataactcgtgtaccttaggttcaaacctcgcactcagtcagtcagaca	6178
Db	6121	accgtgacgaagaataactcgtgtaccttaggttcaaacctcgcactcagtcagtcagaca	6180
Qy	6179	cacgcaggcacacaagccaaaaactctgacgcagcaggtgatacaacccacacagcaggt	6238
Db	6181	caocgagccacacaagccaaaaactctgacgcagcaggtgatacaacccacacagcaggt	6240
Qy	6239	gcaacgggtacggttaaggccttctcgtgacaaaagcggttggtgcgctcccgctgggt	6298
Db	6241	gcaacgggtacggttaaggccttctcgtgacaaaagcggttggtgcgctcccgctgggt	6300
Qy	6299	gcctcagggtgtgaacgcggtatcccaaaatgtggcgagcagtgagggtcagtcgcaaccagc	6358
Db	6301	gcctcagggtgtgaacgcggtatcccaaaatgtggcgagcagtgagggtcagtcgcaaccagc	6360
Qy	6359	acgatcgggtcaatgtgagcaggtgtacaagcccaagcattgcgaacgcgaacc	6418
Db	6361	acgatcgggtcaatgtgagcaggtgtacaagcccaagcattgcgaacgcgaacc	6420
Qy	6419	aatgagcttgaccatcgtatccaccaaaaacgaataaaggccaatgcagggaattcatca	6478
Db	6421	aatgagcttgaccatcgtatccaccaaaaacgaataaaggccaatgcagggaattcatca	6480
Qy	6479	gcgatgcgatggcggtccatgccaagcctcatctcctggcagatccatgggttacccggg	6538
Db	6481	gcgatgcgatggcggtccatgccaagcctcatctcctggcagatccatgggttacccggg	6540
Qy	6539	ggatttgccacccacaacggtcaaggtgcggtggcgatgggactgtcgaagctgtcggtat	6598
Db	6541	ggatttgccacccacaacggtcaaggtgcggtggcgatgggactgtcgaagctgtcggtat	6600
Qy	6599	aatggccaatgggtatttaaaalcaatggttcagccgatacccaaggccatgtagggcg	6658
Db	6601	aatggccaatgggtatttaaaalcaatggttcagccgatacccaaggccatgtagggcg	6660
Qy	6659	gcagttggtgcaggttttccacttttaagccataaaatcgcagaagatttacttaaaaaatcaa	6718
Db	6661	gcagttggtgcaggttttccacttttaagccataaaatcgcagaagatttacttaaaaaatcaa	6720
Qy	6719	tctacacatagttgtat-aaaacagcatcagcatcagtcattactgactgactgattt	6777
Db	6721	tctacacatagttgtat-aaaacagcatcagcatcagtcattactgactgactgattt	6780
Qy	6778	ttataccttaaacacattttaccgctccaagtgattctctcttccacatgacccaatcgcca	6837
Db	6781	ttataccttaaacacattttaccgctccaagtgattctctcttccacatgacccaatcgcca	6840

Qy	6838	ttgatcataggtaaaccttattagtagtaaatatttatcaaatgtagtattgattagataatgggttaaa	6899
Db	6841	TTGATCATAGGTAACCTTATTAGTAAATTTTATCAATGTAGTTGTGTAGATATGGTTAAA	6900
Qy	6898	atttgccattgaacaaaaatgacgatttatcccgaaaaatttctgatttatgatccgtt	6957
Db	6901	ATTGTGCCATTGACCAAAAAAATGACCGATTATCCGAAAAATTTCTGATTTATGATCCGTT	6960
Qy	6958	gacctgcaggtgcag 6972	
Db	6961	GACCTGCAGGTGCAG 6975	
RESULT 12			
US-08-431-718C-1			
; Sequence 1, Application US/08431718C			
; GENERAL INFORMATION:			
; APPLICANT: Ken, Sasaki			
; APPLICANT: Robin, Harkness E.			
; APPLICANT: Loosmore, Sheena M.			
; APPLICANT: Klein, Michel H.			
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN O			
; TITLE OF INVENTION: MORAXELLA			
; FILE REFERENCE: 1038-429 MIS			
; CURRENT APPLICATION NUMBER: US/08/431,718C			
; CURRENT FILING DATE: 1995-05-01			
; NUMBER OF SEQ ID NOS: 4			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 6975			
; TYPE: DNA			
; ORGANISM: Moraxella catarrhalis			
US-08-431-718C-1			
Query Match 99.5%; Score 6939; DB 8; Length 6975;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 6972; Conservative 0; Mismatches 0; Indels 3; Gaps			
Qy	1	ccatggatatgggcagggtgtgcctgcctgcctgatgatgacgcatgacacccatttgccc	60
Db	1	ccaTggatatgggcagggtgTgcctgcctgcctgatgatgacgcatgacacccatttgccc	60
Qy	61	catactgtacgatttgacatgtgatatttaacatgtgcacatgatttaacattgttt	120
Db	61	catactgtacgatttgacatgtgatatttaacatgtgcacatgatttaacattgttt	120
Qy	121	aatactgttgccatccattaccataatttagtgaacgcatatttagtaacgcatattgtataa	180
Db	121	aatactgttgccatccattaccataatttagtgaacgcatatttagtaacgcatattgtataa	180
Qy	181	cattgcgcccctttatgtatcatatgaatagaatattatgatctgattgatttctgtatctgt	240
Db	181	cattgcgcccctttatgtatcatatgaatagaatattatgatctgattgatttctgtatctgt	240
Qy	241	atcagaatgggtgatgctcatatgatgagtcctacgagttgatttgggttaatacactctatg	300
Db	241	atcagaatgggtgatgctcatatgatgagtcctacgagttgatttgggttaatacactctatg	300
Qy	301	atttgatatatttggaaactaatctattgacttaataccatacattggtttataattatgca	360
Db	301	atttgatatatttggaaactaatctctgacttaataccatacattggtttataattatgca	360
Qy	361	taatggtaggctttttgtaaaatcacatgcgcaatatttctactgtttactaccatgct	420
Db	361	taatggtaggctttttgtaaaatcacatgcgcaatatttctactgtttactaccatgct	420
Qy	421	tgaatgcagatcccaatccaccagattcattcaagtagtatgtttgtatagcgaccatttta	480
Db	421	tgaatgcagatcccaatccaccagattcattcaagtagtatgtttgtatagcgaccatttta	480
Qy	481	ccctaattatttcaatcaaatggcctatgtcagcagtgctatcatttttttaagggttaaacac	540

Db 481 cccataattattcaatacaaatgcctatgtcagcgtgtatcatcttttttaagggtataaacac 540
Qy 541 catgaatcaatctataaaagtcattcttaacaaagccacagggcacatttatggcagtgcc 600
Db 541 catgaatcaatctataaaagtcattcttaacaaagccacagggcacatttatggcagtgcc 600
Qy 601 agagtacgcaaaatccacacagcac-gggggggggtagctgtctacagggcacaagtggca 659
Db 601 agagtacgcaaaatccacacagcacgggggggggtagctgtctacagggcacaagtggca 660
Qy 660 gtgtatcaactctgagctttgcccgtattgcgcgctgcgtctgcctcgtatcggtgcaa 719
Db 661 gtgtatcaactctgagctttgcccgtattgcgcgctgcgtctgcctcgtatcggtgcaa 720
Qy 720 cgcctagtgccagtgcttatgctcaaaaaaagataccaaacatatcgcaattggtgaac 779
Db 721 cgcctagtgccagtgcttatgctcaaaaaaagataccaaacatatcgcaattggtgaac 780
Qy 780 aaaaaccagcaagcgtcagcgcaactgccaagggcgagcgtgacgcagccattgctattg 839
Db 781 aaaaaccagcaagcgtcagcgcaactgccaagggcgagcgtgacgcagccattgctattg 840
Qy 840 gtgaaaatgctaaacgcacagggcggtcaagccatcgccatcggttagtagtaataaaaactg 899
Db 841 gtgaaaatgctaaacgcacagggcggtcaagccatcgccatcggttagtagtaataaaaactg 900
Qy 900 tcaatggaagcagtttggaataagtaggtaccggtacggtacggttcaagagtcacatcgcca 959
Db 901 tcaatggaagcagtttggaataagtaggtaccggtacggttcaagagtcacatcgcca 960
Qy 960 tcggtgtgagttaaaagcgtaggtgacgctcgtcgtatgccaatcggttagtgacttac 1019
Db 961 tcggtgtgagttaaaagcgtaggtgacgctcgtcgtatgccaatcggttagtgacttac 1020
Qy 1020 attgctgtgacagctgttaactcctaaacatccgaaggtactctgattaaacgatctta 1079
Db 1021 attgctgtgacagctgttaactcctaaacatccgaaggtactctgattaaacgatctta 1080
Qy 1080 ttaacgcccactgcagttataaaagaataacgaagctcaaaaggataatgatgtaaaata 1139
Db 1081 ttaacgcccactgcagttataaaagaataacgaagctcaaaaggataatgatgtaaaata 1140
Qy 1140 gacgcacaacgcgaagcggacacgcagctgactgcagtgaggagccaatgcatatgcacagg 1199
Db 1141 gacgcacaacgcgaagcggacacgcagctgactgcagtgaggagccaatgcatatgcacagg 1200
Qy 1200 gtcaattttccaaacgctttgtgtacacgggccaacagctaaaaagtgcctattccctggcag 1259
Db 1201 gtcaattttccaaacgctttgtgtacacgggccaacagctaaaaagtgcctattccctggcag 1260
Qy 1260 tgggtcttgccgcacagccagcgggccaatctacaatctcgtattggttctgtagcaacat 1319
Db 1261 tgggtcttgccgcacagccagcgggccaatctacaatctcgtattggttctgtagcaacat 1320
Qy 1320 ctactcgttgggagcagatagcccttggtcaggtactcgtgctcagctacagggcagta 1379
Db 1321 ctactcgttgggagcagatagcccttggtcaggtactcgtgctcagctacagggcagta 1380
Qy 1380 ttgcccctaggttcaagggtctgtgttgcactcagagtgataataatctctagaccggccctata 1439
Db 1381 ttgcccctaggttcaagggtctgtgttgcactcagagtgataataatctctagaccggccctata 1440
Qy 1440 caccaaataccagccagcactagaccccaagtttcaagccacccaataataacgaagcgggtc 1499
Db 1441 caccaaataccagccagcactagaccccaagtttcaagccacccaataataacgaagcgggtc 1500
Qy 1500 cactttccattggttagtaactctataaaagtaaaaaatcataatgctcgggtcaggtgtta 1559
Db 1501 cactttccattggttagtaactctataaaagtaaaaaatcataatgctcgggtcaggtgtta 1560
Qy 1560 ataaaaccgctgcggttcaatgtggtcacagctagaaagcgggtggtgaagtgggttaagagc 1619
Db 1561 ataaaaccgctgcggttcaatgtggtgacagctagaaagcgggtggtgaagtgggttaagagc 1620

Qy 1620 gtagaattacttttcadgggtgatataaacagctactgacgtataaaataggttttgataata 1679
Db 1621 gtagaattacttttcadgggtgatataaacagctactgacgtataaaataggttttgataata 1680
Qy 1680 ctttaactatttaaagggtggtgcagagacccaacgcattaaaccgataataatcctggtg 1739
Db 1681 ctttaactatttaaagggtggtgcagagacccaacgcattaaaccgataataatcctggtg 1740
Qy 1740 taaaagaggtgataaagtggtctgaaagttaaacttgctaaaactttaaacatctta 1799
Db 1741 taaaagaggtgataaagtggtctgaaagttaaacttgctaaaactttaaacatctta 1800
Qy 1800 ctgaggtgaaatcacaaactacattaaatgccacaacacagtttaaggtaggtagtagtagta 1859
Db 1801 ctgaggtgaaatcacaaactacattaaatgccacaacacagtttaaggtaggtagtagtagta 1860
Qy 1860 gtactacagctgaattatttgatgtagtattaaacctttaccagcccaatcacaggcagtc 1919
Db 1861 gtactacagctgaattatttgatgtagtattaaacctttaccagcccaatcacaggcagtc 1920
Qy 1920 aaagcacaagcaaaacccgtctatggcgttaatgggtgaaagttaacttaataatgcagaaa 1979
Db 1921 aaagcacaagcaaaacccgtctatggcgttaatgggtgaaagttaacttaataatgcagaaa 1980
Qy 1980 caacagcagcaaatcggcactactcgtattaccagagataaaaattggctttgctcgagatg 2039
Db 1981 caacagcagcaaatcggcactactcgtattaccagagataaaaattggctttgctcgagatg 2040
Qy 2040 gtgagtgttgataaaaaacaacagcacacataatttggtataaaaaaacacttaaaagtggtagtg 2099
Db 2041 gtgagtgttgataaaaaacaacagcacacataatttggtataaaaaaacacttaaaagtggtagtg 2100
Qy 2100 ttgcaattaccatagacaatggcattgatcaggtataaaaaaatcagtaaatcttgcca 2159
Db 2101 ttgcaattaccatagacaatggcattgatcaggtataaaaaaatcagtaaatcttgcca 2160
Qy 2160 aaggtagcagtgctaaacgagtcggtttaccatcgcaacagctcaaaagccgcccagcctactt 2219
Db 2161 aaggtagcagtgctaaacgagtcggtttaccatcgcaacagctcaaaagccgcccagcctactt 2220
Qy 2220 taaacgcagggcgtggtcactcaggtgcacacactactgaaataatcagttgactgaagatg 2279
Db 2221 taaacgcagggcgtggtcactcaggtgcacacactactgaaataatcagttgactgaagatg 2280
Qy 2280 gcaatgtttaccgccccaaacttacaacattggcgtgaaaaaccacccagcgttaacagtgatg 2339
Db 2281 gcaatgtttaccgccccaaacttacaacattggcgtgaaaaaccacccagcgttaacagtgatg 2340
Qy 2340 gcaatgtgataaaatttagtgtttaagggttagtgtacgaacaataatgcttagttaccgccc 2399
Db 2341 gcaatgtgataaaatttagtgtttaagggttagtgtacgaacaataatgcttagttaccgccc 2400
Qy 2400 aacatttggcgaagctatataaataagtcacacacggtgacaggtgctctacaaagct 2459
Db 2401 aacatttggcgaagctatataaataagtcacacacggtgacaggtgctctacaaagct 2460
Qy 2460 ttaccgtttaaagaagaagacgatgatgacgcacacgctcaccgtggtctaaaagatacga 2519
Db 2461 ttaccgtttaaagaagaagacgatgatgacgcacacgctcaccgtggtctaaaagatacga 2520
Qy 2520 caaaaaatgcccgcgcagtcagtcattttaaactcaaaaggttaaaacggttctaacggttg 2579
Db 2521 caaaaaatgcccgcgcagtcagtcattttaaactcaaaaggttaaaacggttctaacggttg 2580
Qy 2580 ctacccaaaaaagatgggtacggtttacccttgggttagccaagatagcgttctgaccttg 2639
Db 2581 ctacccaaaaaagatgggtacggtttacccttgggttagccaagatagcgttctgaccttg 2640
Qy 2640 gcaaaagcacccttaaaacaacgatgcttgactgttaagatataccaacgacaaatcaag 2699
Db 2641 gcaaaagcacccttaaaacaacgatgcttgactgttaagatataccaacgacaaatcaag 2700

Db 4861 ctcaaatgaatgtcaaaatcagtcattacaagaagacaagtaaatgatgccaaataaaaagc 4920
Qy 4920 aaggcatcaatgaagacacgcctttgttaaagacattgaaagacgccttcttgataaca 4979
Db 4921 aaggcatcaatgaagacacgcctttgttaaagacattgaaagacgccttcttgataaca 4980
Qy 4980 aaacccaaaaacgcgcagtaactgtgggtgattttaaattgcggttgcccaaacacgcgtga 5039
Db 4981 aaacccaaaaacgcgcagtaactgtgggtgattttaaattgcggttgcccaaacacgcgtga 5040
Qy 5040 cctttgcagggtatcacaggcacaacgcgttaaaaaactggcgagacatttgaccatcaaaag 5099
Db 5041 cctttgcagggtatcacaggcacaacgcgttaaaaaactggcgagacatttgaccatcaaaag 5100
Qy 5100 gtgggcaacacagacacccaataagctaaaccgataataacatcggtgtgtgtagcaggtactg 5159
Db 5101 gtgggcaacacagacacccaataagctaaaccgataataacatcggtgtgtgtagcaggtactg 5160
Qy 5160 atggcttcaactgtcaaaacttgccaaagaccttaaccaatctttaaagcgtttaatgcaggtg 5219
Db 5161 atggcttcaactgtcaaaacttgccaaagaccttaaccaatctttaaagcgtttaatgcaggtg 5220
Qy 5220 gcaccaaaattgatgacaaaagcggtgtctttgttagactcaagcggtcaagcccaaaagcaa 5279
Db 5221 gcaccaaaattgatgacaaaagcggtgtctttgttagactcaagcggtcaagcccaaaagcaa 5280
Qy 5280 acacccctgtctaaagtccaatggcctggacctgggtggcaaggtcatcagtaattgtgg 5339
Db 5281 acacccctgtctaaagtccaatggcctggacctgggtggcaaggtcatcagtaattgtgg 5340
Qy 5340 gcaaaaggcacaaaagataccga - cgtctgccaattgtacaacagtttaaacgaagtagcgaac 5398
Db 5341 gcaaaaggcacaaaagataccgaagcgtcccaattgtacacagatttaaacgaagtagcgaac 5400
Qy 5399 ttgttgggtctgttaattgctgtaataacgcgtgcagcgaatcaggttaaaacattgcc 5458
Db 5401 ttgttgggtctgttaattgctgtaataacgcgtgcagcgaatcaggttaaaacattgcc 5460
Qy 5459 gacatcaaaaagaccacaaatcaggttcatatcatatcaatcccaactgtcatcaaaagcagcc 5518
Db 5461 gacatcaaaaagaccacaaatcaggttcatatcatatcaatcccaactgtcatcaaaagcagcc 5520
Qy 5519 acggtacttggcgttaaaagttaaataacgataccgaaatacttgcacattggtgggtatacaa 5578
Db 5521 acggtacttggcgttaaaagttaaataacgataccgaaatacttgcacattggtgggtatacaa 5580
Qy 5579 gtggcggtggataaagacggaacgctaacgcgatttaagcaatgtttgggtcaaaaacc 5638
Db 5581 gtggcggtggataaagacggaacgctaacgcgatttaagcaatgtttgggtcaaaaacc 5640
Qy 5639 caaaaagatggcagcaaaaagccctgctgccactttaacgcgcgcggttcagaccacac 5698
Db 5641 caaaaagatggcagcaaaaagccctgctgccactttaacgcgcggttcagaccacac 5700
Qy 5699 tatttgacaaacacccgcgagaagccaattgacagaaataatgaaacaaaggtatccgcttc 5758
Db 5701 tatttgacaaacacccgcgagaagccaattgacagaaataatgaaacaaaggtatccgcttc 5760
Qy 5759 ttccatgtcaacgatggcaataagagcctgtgtgtaacagggcgttaacgcgcatgtactca 5818
Db 5761 ttccatgtcaacgatggcaataagagcctgtgtgtaacagggcgttaacgcgcatgtactca 5820
Qy 5819 agtgcctcaggcaagcactcagtgccgataggtttccaggccaaagcagatgggtgaagcc 5878
Db 5821 agtgcctcaggcaagcactcagtgccgataggtttccaggccaaagcagatgggtgaagcc 5880
Qy 5879 gccgttgcatagggagacaaacccaacgagcgaacccaatccatccgcatcgggtgataac 5938
Db 5881 gccgttgcatagggagacaaacccaacgagcgaacccaatccatccgcatcgggtgataac 5940
Qy 5939 gcacaagccacggcgatcaatccatccatcggtacagcaaatgtgtgtagcaggttaag 5998
Db 5941 gcacaagccacggcgatcaatccatccatcggtacagcaaatgtgtgtagcaggttaag 6000

Qy 5999 cactctgtgccaatgcgcgaacccaagcactgttaaggctgataacagttacagttggtggt 6058
Db 6001 cactctgtgccaatgcgcgaacccaagcactgttaaggctgataacagttacagttggtggt 6060
Qy 6059 aataacaacaacagtttacccgatgccactcaaacccagctcttcttgggtgggcaataaacatc 6118
Db 6061 aataacaacaacagtttacccgatgccactcaaacccagctcttcttgggtgggcaataaacatc 6120
Qy 6119 accgtgacccgaaagttaactcggttgcttaggttcaaaactctgccatcagtcagggcaca 6178
Db 6121 accgtgacccgaaagttaactcggttgcttaggttcaaaactctgccatcagtcagggcaca 6180
Qy 6179 cacgcaggcacacaagcccaaaaatactgacggcacagcaggttacaccaccacacagcaggt 6238
Db 6181 cacgcaggcacacaagcccaaaaatactgacggcacagcaggttacaccaccacacagcaggt 6240
Qy 6239 gcaacccgttaacggttaaaaggtttgtgtggaacaaacgcggttgggtgcggtctccgtggt 6298
Db 6241 gcaacccgttaacggttaaaaggtttgtgtggaacaaacgcggttgggtgcggtctccgtggt 6300
Qy 6299 gcctcaggtgctgaacgccttatccaaaatgtgagcagcaggtgaggtcagtcgccaccagc 6358
Db 6301 gcctcaggtgctgaacgccttatccaaaatgtgagcagcaggtgaggtcagtcgccaccagc 6360
Qy 6359 accgatgcggtcgaatggtagccagttgtacaaaagccacccaaagcattgccaacgcaacc 6418
Db 6361 accgatgcggtcgaatggtagccagttgtacaaaagccacccaaagcattgccaacgcaacc 6420
Qy 6419 aatgagcttgaccatcgctatccacaaacgaaataaaggccaatgaggggatttcaaca 6478
Db 6421 aatgagcttgaccatcgctatccacaaacgaaataaaggccaatgaggggatttcaaca 6480
Qy 6479 gcgatggcgatggcggtccatgcacaaacgctcacatctctgcagatccatggttaccggg 6538
Db 6481 gcgatggcgatggcggtccatgcacaaacgctcacatctctgcagatccatggttaccggg 6540
Qy 6539 ggtattgcccaccccaacacggttcaaggtgcggtggcagtgaggactgctgaagctgtcgat 6598
Db 6541 ggtattgcccaccccaacacggttcaaggtgcggtggcagtgaggactgctgaagctgtcgat 6600
Qy 6599 aatggtcaatgggtatttataaatacaatggttcagcgatatacccaaggccatgtagggcg 6658
Db 6601 aatggtcaatgggtatttataaatacaatggttcagcgatatacccaaggccatgtagggcg 6660
Qy 6659 gcagttggtgcaggttttcaacttttaagccataaatacgaagattttacttaaaaatacaa 6718
Db 6661 gcagttggtgcaggttttcaacttttaagccataaatacgaagattttacttaaaaatacaa 6720
Qy 6719 tctcaccatagttgtat - aaaaacgatacagcatcagcatcactatattactgatgtatgttt 6777
Db 6721 tctcaccatagttgtataaaaacgatacagcatcagcatcactatattactgatgtatgttt 6780
Qy 6778 tttatcaacttaaacattttaccgctcaagtgattctctttccacatgaccaaaatcgcca 6837
Db 6781 tttatcaacttaaacattttaccgctcaagtgattctctttccacatgaccaaaatcgcca 6840
Qy 6838 ttgatcataggttaaaacttattgagtaaattttatcaatgtagttgttagatatgtgtaaa 6897
Db 6841 ttgatcataggttaaaacttattgagtaaattttatcaatgtagttgttagatatgtgtaaa 6900
Qy 6898 atttgccattgacaaaaaaatgaccgattttatcccgaaaaatttctgattatgacccgtt 6957
Db 6901 atttgccattgacaaaaaaatgaccgattttatcccgaaaaatttctgattatgacccgtt 6960
Qy 6958 gacctgacgggtcgac 6972
Db 6961 gacctgacgggtcgac 6975

RESULT 13

US-09-361-619-12

; Sequence 12, Application US/09361619

708	tgatcgtgcaacgcgtcagtgccagtgctta	tgctcaaaaaaaagataccaaaacata	tcg	767
2	tgatcgtgcaacgcgtcagtgccagtgctta	tgctcaaaaaaaagataccaaaacata	tcg	61
	tgatcgtgcaacgcgtcagtgccagtgctta	tgctcaaaaaaaagataccaaaacata	tcg	61
768	caattggtagaaaaaacccagccaagaacgcgtc	caggcactgccaaagcgagcgtgatcgaag	827	
62	caattggtagaaaaaacccagccaagaacgcgtc	caggcactgccaaagcgagcgtgatcgaag	121	
	caattggtagaaaaaacccagccaagaacgcgtc	caggcactgccaaagcgagcgtgatcgaag	121	
828	ccattgctattgtagaaatgctaaacgcacagcg	cggctcaagccatcgccatcgcgtagtga	887	
122	ccattgctattgtagaaatgctaaacgcacagcg	cggctcaagccatcgccatcgcgtagtga	181	
	ccattgctattgtagaaatgctaaacgcacagcg	cggctcaagccatcgccatcgcgtagtga	181	
888	gtaataaaactgtcaatgggaagcagttggata	agaagtaggtaccgatgctacgggtccaag	947	
182	gtaataaaactgtcaatgggaagcagttggata	agaagtaggtaccgatgctacgggtccaag	241	
	gtaataaaactgtcaatgggaagcagttggata	agaagtaggtaccgatgctacgggtccaag	241	
948	agtcacatcgccatcggtgggtgtagtaaaagg	ctagtggtgatccctcgattgccatcggtga	1007	
242	agtcacatcgccatcggtgggtgtagtaaaagg	ctagtggtgatccctcgattgccatcggtga	301	
	agtcacatcgccatcggtgggtgtagtaaaagg	ctagtggtgatccctcgattgccatcggtga	301	
1008	gtgatgacttacattgtcttgatcagcatggt	taactctaaacatcgaaaggtactctga	1067	
302	gtgatgacttacattgtcttgatcagcatggt	taactctaaacatcgaaaggtactctga	361	
	gtgatgacttacattgtcttgatcagcatggt	taactctaaacatcgaaaggtactctga	361	
1068	ttaacgatctttattaaacggccatgcagtag	tataaaagaatacacgaagctcaaaagataa	1127	
362	ttaacgatctttattaaacggccatgcagtag	tataaaagaatacacgaagctcaaaagataa	421	
	ttaacgatctttattaaacggccatgcagtag	tataaaagaatacacgaagctcaaaagataa	421	
1128	atgtaaaaataagacgcacaaacccaagcgga	cacacgcagtagctcagtgggagccatgt	1187	
422	atgtaaaaataagacgcacaaacccaagcgga	cacacgcagtagctcagtgggagccatgt	481	
	atgtaaaaataagacgcacaaacccaagcgga	cacacgcagtagctcagtgggagccatgt	481	
1188	catatgcacaggggtcatttttccaaacgcctt	tggtacacgggcaacagctaaaagtgcct	1247	
482	catatgcacaggggtcatttttccaaacgcctt	tggtacacgggcaacagctaaaagtgcct	541	
	catatgcacaggggtcatttttccaaacgcctt	tggtacacgggcaacagctaaaagtgcct	541	
1248	attccttggcagtggtggtcttgcccgccacag	cgcagggccaattcacaaatcgctatgggt	1307	
542	attccttggcagtggtggtggtcttgcccgccac	agggccaattcacaaatcgctatgggt	601	
	attccttggcagtggtggtggtggtcttgcccgcc	agggccaattcacaaatcgctatgggt	601	
1308	ctgatagcacaactagctcgttggagcgtatag	cccttggctgaggttactcgtgcctcagc	1367	
602	ctgatagcacaactagctcgttggagcgtatag	cccttggctgaggttactcgtgcctcagc	661	
	ctgatagcacaactagctcgttggagcgtatag	cccttggctgaggttactcgtgcctcagc	661	
1368	tacaggcgagttatgcccctaggtcaaggttct	gtgtgtcactcagagtgataaataattcta	1427	
662	tacaggcgagttatgcccctaggtcaaggttct	gtgtgtcactcagagtgataaataattcta	721	
	tacaggcgagttatgcccctaggtcaaggttct	gtgtgtcactcagagtgataaataattcta	721	
1428	gaccggccttatcacccaataaccaggcaact	agaccacgaatttcaagcgaccacaata	1487	

Qy 2568 gtctaacggttgctaccacaaagagatggttaacgtttacaccttgggcttagccaaagatagcg 2627
Db 1862 gtctaacggttgctaccacaaagagatggttaacgtttacaccttgggcttagccaaagatagcg 1921
Qy 2628 gtctgaccattggcaaaagcaccctaaacacacgctgcttgactgtttaagatcaccaacg 2687
Db 1922 gtctgaccattggcaaaagcaccctaaacacacgctgcttgactgtttaagatcaccaacg 1981
Qy 2688 aacaaatccaagtcggtgcttaattgacattaaatttacttaattgaaatggtagtaattccag 2747
Db 1982 aacaaatccaagtcggtgcttaattgacattaaatttacttaattgaaatggtagtaattccag 2041
Qy 2748 gtaactgcatgcaaatcacgctgcgtactaccagagataaaattggcttggctggtctg 2807
Db 2042 gtaactgcatgcaaatcacgctgcgtactaccagagataaaattggcttggctggtctg 2101
Qy 2808 atggtgcagttgtacaaacaaaccttacttgcatacgaacaaagctacaagttggcaatg 2867
Db 2102 atggtgcagttgtacaaacaaaccttacttgcatacgaacaaagctacaagttggcaatg 2161
Qy 2868 ttaagattaccaacacgtggttaacgcaggtggttaagccatcacagggctgtccccaa 2927
Db 2162 ttaagattaccaacacgtggttaacgcaggtggttaagccatcacagggctgtccccaa 2221
Qy 2928 cactgcttagcattgcccagatcaaaagtgcgcgaacatagaaactgggcaatacaatccaa 2987
Db 2222 cactgcttagcattgcccagatcaaaagtgcgcgaacatagaaactgggcaatacaatccaa 2281
Qy 2988 acaagacaaatccaaacgctgcagcagcattaatgatatattaaatacagggctttaacctaa 3047
Db 2282 acaagacaaatccaaacgctgcagcagcattaatgatatattaaatacagggctttaacctaa 2341
Qy 3048 aaaataatacaaccccatgactttgtctccacttatgacattgttgactttgccaatg 3107
Db 2342 aaaataatacaaccccatgactttgtctccacttatgacattgttgactttgccaatg 2401
Qy 3108 gcaatgcaccacgcgcacagtaaccatgataccgcttaacaaacacagtaaaagtgtat 3167
Db 2402 gcaatgcaccacgcgcacagtaaccatgataccgcttaacaaacacagtaaaagtgtat 2461
Qy 3168 atgattgaaatgtgattgatacaacacattcatctaaacaggaactgatacaataaaaaac 3227
Db 2462 atgattgaaatgtgattgatacaacacattcatctaaacaggaactgatacaataaaaaac 2521
Qy 3228 ttggcgtcaaaacccacaaactgacaaacaaagtgctaaatggtaatacagcaactaaat 3287
Db 2522 ttggcgtcaaaacccacaaactgacaaacaaagtgctaaatggtaatacagcaactaaat 2581
Qy 3288 ttaatgttaactctagtgaagaatgccctgtttaacgccaagagacatgcgcgaataatc 3347
Db 2582 ttaatgttaactctagtgaagaatgccctgtttaacgccaagagacatgcgcgaataatc 2641
Qy 3348 taaacccctagcacaagaaattcacaccacaaagcagacagacacgcgcctacaaa 3407
Db 2642 taaacccctagcacaagaaattcacaccacaaagcagacagacacgcgcctacaaa 2701
Qy 3408 cctttacggttaaaaaggttagatgaaataataatgctgtagcgcgaagcgcacacccg 3467
Db 2702 cctttacggttaaaaaggttagatgaaataataatgctgtagcgcgaagcgcacacccg 2761
Qy 3468 tgggtcaaaaagacgcaataataacgttcaacacccctaaacactcaaaaggtgaaacggctc 3527
Db 2762 tgggtcaaaaagacgcaataataacgttcaacacccctaaacactcaaaaggtgaaacggctc 2821
Qy 3528 ttaataataaacacgcaaaaatggttagcttaccttggcattaaacacacacagcggtc 3587
Db 2822 ttaataataaacacgcaaaaatggttagcttaccttggcattaaacacacacagcggtc 2881
Qy 3588 ttaagcgcggcaaaagcaccctaaacgacgggtgcttgcattataaaaccccaactggta 3647
Db 2882 ttaagcgcggcaaaagcaccctaaacgacgggtgcttgcattataaaaccccaactggta 2941

Qy 3648 gcgaacaaatccaagtcggtgctgctgaggtggaagtttgccaaaggttaataataatggtg 3707
Db 2942 gcgaacaaatccaagtcggtgctgctgaggtggaagtttgccaaaggttaataataatggtg 3001
Qy 3708 ttgtagatgctggcattgattgagcacaactcgattaccagagatgaaattggcttactg 3767
Db 3002 ttgtagatgctggcattgattgagcacaactcgattaccagagatgaaattggcttactg 3061
Qy 3768 ggactaatggctcacttgataaaagcaaaccccccactaaagcaaaagcgcattaaacgcag 3827
Db 3062 ggactaatggctcacttgataaaagcaaaccccccactaaagcaaaagcgcattaaacgcag 3121
Qy 3828 gtgtaaaagatgatacaacattcaatcaagtgagattgcccacaaacgcgcattgctg 3887
Db 3122 gtgtaaaagatgatacaacattcaatcaagtgagattgcccacaaacgcgcattgctg 3181
Qy 3888 tgacagcggcgaagatttatgtttaaaacsgaacctgtaaaacacaaatcagcagtaactg 3947
Db 3182 tgacagcggcgaagatttatgtttaaaacsgaacctgtaaaacacaaatcagcagtaactg 3241
Qy 3948 ccaaaacagcacaaaactcattcacgaattctcagtagcagatgaacaaaggttaataact 4007
Db 3242 ccaaaacagcacaaaactcattcacgaattctcagtagcagatgaacaaaggttaataact 3301
Qy 4008 ttacggttagtaaccccttactccagttatgacacctcaaaagacctctgactacact 4067
Db 3302 ttacggttagtaaccccttactccagttatgacacctcaaaagacctctgactacact 3361
Qy 4068 ttgcagatgaaacgcgcattaccacaaaggtataataaaggttggtgcttgaggcattg 4127
Db 3362 ttgcagatgaaacgcgcattaccacaaaggtataataaaggttggtgcttgaggcattg 3421
Qy 4128 accaaacaaaggttaaccacgcctaaagtcgcgtgggttaataaaggttggtgcttgaggcattg 4187
Db 3422 accaaacaaaggttaaccacgcctaaagtcgcgtgggttaataaaggttggtgcttgaggcattg 3481
Qy 4188 ttgtcattgacagccaaatggttcaaaataccatcaacaggaactgaagcaactctagcta 4247
Db 3482 ttgtcattgacagccaaatggttcaaaataccatcaacaggaactgaagcaactctagcta 3541
Qy 4248 atgttacaatgataaaggttagctagcaccacagacagggcaataataataaagacg 4307
Db 3542 atgttacaatgataaaggttagctagcaccacagacagggcaataataataaagacg 3601
Qy 4308 aagcaaaacccgctgcgcgcagcattgttgatgcttaagcgcaggttttaacttgcaag 4367
Db 3602 aagcaaaacccgctgcgcgcagcattgttgatgcttaagcgcaggttttaacttgcaag 3661
Qy 4368 gcaatggtgaagcgggttgactttgtctccacttatgacacgcgtcaactttgcccagtgca 4427
Db 3662 gcaatggtgaagcgggttgactttgtctccacttatgacacgcgtcaactttgcccagtgca 3721
Qy 4428 atgcccacccgctaaaggtgacctatgatacaagcaaaacacagtaaaagtgtctatg 4487
Db 3722 atgcccacccgctaaaggtgacctatgatacaagcaaaacacagtaaaagtgtctatg 3781
Qy 4488 atgtcaatgttgatgatacaacaccttgaagttaaagaataaaacttgccgttaaaaccca 4547
Db 3782 atgtcaatgttgatgatacaacaccttgaagttaaagaataaaacttgccgttaaaaccca 3841
Qy 4548 ccaattgaccagttactggcaggtgctataaaattggcctgaagcaactcaagctactg 4607
Db 3842 ccaattgaccagttactggcaggtgctataaaattggcctgaagcaactcaagctactg 3901
Qy 4608 gcgatgcgcttgcagacagcagtgatctgctcattcaaacacacttatctgcgcaca 4667
Db 3902 gcgatgcgcttgcagacagcagtgatctgctcattcaaacacacttatctgcgcaca 3961
Qy 4668 tccaaactgcaaaagggcgaagcgaacactcagcaggttatgtggtgctgctg 4727
Db 3962 tccaaactgcaaaagggcgaagcgaacactcagcaggttatgtggtgctgctg 4021
Qy 4728 gcaataaggttcatctatgacagtgaccgataacaaagtactatcaagccaaaatgatggca 4787

|||||
Db 4022 gcaataaggccaactctatgacagcacagataacagactactatcaagccaaaaatgtaggca 4081
QY 4788 cagttgataaaacccaaagaagttgccaagacacaaactggtcgcccaagcccaaaccccccag 4847
Db 4082 cagttgataaaacccaaagaagttgccaagacacaaactggtcgcccaagcccaaaccccccag 4141
QY 4848 atggcacattggctcctaaatgaaatgtaaatcagtcattcaataaagaacaaagtaaatgtag 4907
Db 4142 atggcacattggctcctaaatgaaatgtaaatcagtcattcaataaagaacaaagtaaatgtag 4201
QY 4908 ccaataaaagcagacatcaatgaaagacacacgcccctttgttaagagacttgaaaaagccg 4967
Db 4202 ccaataaaagcagacatcaatgaaagacacacgcccctttgttaagagacttgaaaaagccg 4261
QY 4968 ctctctgataaaacccaaacccgacgagtaactggtggtgatttaaatgcccgttgccc 5027
Db 4262 ctctctgataaaacccaaacccgacgagtaactggtggtgatttaaatgcccgttgccc 4321
QY 5028 aaacacgcgtgacacttgacgggatacagggacacacggtacaaaacttgggcgagactt 5087
Db 4322 aaacacgcgtgacacttgacgggatacagggacacacggtacaaaacttgggcgagactt 4381
QY 5088 tgaccatcaaaagtggaacacagacacacaaataagctaacggataataacatcggtgtgg 5147
Db 4382 tgaccatcaaaagtggaacacagacacacaaataagctaacggataataacatcggtgtgg 4441
QY 5148 tagcaggtactgactgcttcaactgtcaaaacttgccaaagacacttaacacagcg 5207
Db 4442 tagcaggtactgactgcttcaactgtcaaaacttgccaaagacacttaacacagcg 4501
QY 5208 ttaatgcaggtggcacacaaataatgtagcaaaaggcggtctctttgtagactcaagcggtc 5267
Db 4502 ttaatgcaggtggcacacaaataatgtagcaaaaggcggtctctttgtagactcaagcggtc 4561
QY 5268 aagccaaagcaaacacccctgtgctaaagtgcacaaaggcggtacactgggtgcaaggtca 5327
Db 4562 aagccaaagcaaacacccctgtgctaaagtgcacaaaggcggtacactgggtgcaaggtca 4621
QY 5328 tcaagtaattggggcaaaagccacaaagataccagacgctgcccattgtacacacagttaaacg 5387
Db 4622 tcaagtaattggggcaaaagccacaaagataccagacgctgcccattgtacacacagttaaacg 4681
QY 5388 aagtcagcaactttggtctgttggtaactggttaagtataacgctgacgcaatcaag 5447
Db 4682 aagtcagcaactttggtctgttggtaactggttaagtataacgctgacgcaatcaag 4741
QY 5448 taacattgcccagacataaaagacacccaaattcaggttcatcatctaaacccgactgtca 5507
Db 4742 taacattgcccagacataaaagacacccaaattcaggttcatcatctaaacccgactgtca 4801
QY 5508 tcaagcagcagcaggtacttgccggtaagggttaataacgataccgaaaaacttgccactg 5567
Db 4802 tcaagcagcagcaggtacttgccggtaagggttaataacgataccgaaaaacttgccactg 4861
QY 5568 gtgggtatacaagtgggcgtggataaagacgccaacgctaacgcggtatttaagcaatgttt 5627
Db 4862 gtgggtatacaagtgggcgtggataaagacgccaacgctaacgcggtatttaagcaatgttt 4921
QY 5628 ggggtcaaaacccaaagatggcagcaaaaagccctgctgccactataacgcgcgag 5687
Db 4922 ggggtcaaaacccaaagatggcagcaaaaagccctgctgccactataacgcgcgag 4981
QY 5688 gtccagccaactatttgacacaaaccccgacagaagccattgacagaataaatgaacaaag 5747
Db 4982 gtccagccaactatttgacacaaaccccgacagaagccattgacagaataaatgaacaaag 5041
QY 5748 gtatccgctcttccatgtcaacgattggcaatcaagagcctgtgtacaagggcgtaacg 5807
Db 5042 gtatccgctcttccatgtcaacgattggcaatcaagagcctgtgtacaagggcgtaacg 5101
QY 5808 gcattgactcaagtgccaggaagcactcagtggtggtggtggttccagggccaagggag 5867
|||||

Db 5102 gaattgactcaagtgccctcaggcaagcactcagtggtgtagatttccaggccaagggag 5161
QY 5868 atggtgaagccgcgcttgccatagggcagacaaacccaagcaggaacacatccatcgcca 5927
Db 5162 atggtgaagccgcgcttgccatagggcagacaaacccaagcaggaacacatccatcgcca 5221
QY 5928 tcggtgataaacgcacaaagccacggcgatcgaatccatcgccatcggtcagggcaatgtgg 5987
Db 5222 tcggtgataaacgcacaaagccacggcgatcgaatccatcgccatcggtcagggcaatgtgg 5281
QY 5988 tagcaggtgaagcactctggtgccatcggtgcagcccaagcactgtaagggtgataaacggt 6047
Db 5282 tagcaggtgaagcactctggtgccatcggtgcagcccaagcactgtaagggtgataaacggt 5341
QY 6048 acagtggtgggttaataacaaacccggttacgcgtgcccactcaaacccggtctcttgggtgg 6107
Db 5342 acagtggtgggttaataacaaacccggttacgcgtgcccactcaaacccggtctcttgggtgg 5401
QY 6108 gcaataacatcacccgtgacggaaagttaactcgttgccttaggttcaaaactctgccatca 6167
Db 5402 gcaataacatcacccgtgacggaaagttaactcgttgccttaggttcaaaactctgccatca 5461
QY 6168 gtgcaggcacacacgcagggcacacaaagccaaaaactctgacggcagcaggtacaacca 6227
Db 5462 gtgcaggcacacacgcagggcacacaaagccaaaaactctgacggcagcaggtacaacca 5521
QY 6228 ccacagcaggtgcaacccgttacgggttaaaaggctttgtggaacaaacggcggttgggtgg 6287
Db 5522 ccacagcaggtgcaacccgttacgggttaaaaggctttgtggaacaaacggcggttgggtgg 5581
QY 6288 tctcogtgggtgcctcaggtgctgaacgcgttatccaaaatgtggcagcaggtgaggtca 6347
Db 5582 tctcogtgggtgcctcaggtgctgaacgcgttatccaaaatgtggcagcaggtgaggtca 5641
QY 6348 gtgccacagcagcaggtgctgcaatggtgagcaggtgtagccaggtgtacaaagccaccccaagcattg 6407
Db 5642 gtgccacagcagcaggtgctgcaatggtgagcaggtgtagccaggtgtacaaagccaccccaagcattg 5701
QY 6408 ccaacgcacacaaatgagcttgaccatcggtatccacaaacgaaaaataaggccaaatggcg 6467
Db 5702 ccaacgcacacaaatgagcttgaccatcggtatccacaaacgaaaaataaggccaaatggcg 5761
QY 6468 ggatttcatcagcagtggtgagtggtgcgttccatgccacaagccctacattcctgagagatca 6527
Db 5762 ggatttcatcagcagtggtgagtggtgcgttccatgccacaagccctacattcctgagagatca 5821
QY 6528 tgggtacccgggggtattgccaccccaacggtcgaagggtcggtggcagtggtgagctgtcga 6587
Db 5822 tgggtacccgggggtattgccaccccaacggtcgaagggtcggtggcagtggtgagctgtcga 5881
QY 6588 agctgtcgggataatggtcgaatgggtatttaaaatcaatgggttcagccgatacccaagggc 6647
Db 5882 agctgtcgggataatggtcgaatgggtatttaaaatcaatgggttcagccgatacccaagggc 5941
QY 6648 atgtaggggcggtggtgaggtgttctacacttttaagccataaatcgcaagattttac 6707
Db 5942 atgtaggggcggtggtgaggtgttctacacttttaagccataaatcgcaagattttac 6001
QY 6708 ttaaaaaataactctcacacatagttgtataaaaaagcagcatcagtcataattactgat 6767
Db 6002 ttaaaaaataactctcacacatagttgtataaaaaagcagcatcagtcataattactgat 6061
QY 6768 gctgatgttttttatcacttaaacacttttaccgctcaagtgattctctttccaccatgac 6827
Db 6062 gctgatgttttttatcacttaaacacttttaccgctcaagtgattctctttccaccatgac 6121
QY 6828 caaatcgccattgatacaggttaaaacttattgagtaaattttatcaatgtagttgtaga 6887
Db 6122 caaatcgccattgatacaggttaaaacttattgagtaaattttatcaatgtagttgtaga 6181
QY 6888 tatgttataaaaatgtgtccattgacacaaaaaaatgacgcgatttatcccgaaaaatctctgatt 6947
Db 6182 tatgttataaaaatgtgtccattgacacaaaaaaatgacgcgatttatcccgaaaaatctctgatt 6241

.....

Qy	4442	aagtgacctatgatgacacagcaaacaccagtaaaagtggtctctatgatgtccaatgtggat	4501
Db	3901	aagtgacctatgatgacacagcaaacaccagtaaaagtggtctctatgatgtccaatgtggat	3960
Qy	4502	gatacaacctatgaagttaaaataaaaaacttgcgctaaaaaacccacacatgaccagt	4561
Db	3961	gatacaacctatgaagttaaaataaaaaacttgcgctaaaaaacccacacatgaccagt	4020
Qy	4562	actggcacaggctgctaataaaattggccttaagcaatcaagctactgcgcatgtggcttgct	4621
Db	4021	actggcacaggctgctaataaaattggccttaagcaatcaagctactgcgcatgtggcttgct	4080
Qy	4622	aaggccagtgatatcgctgtctcatctaaacaccttatctggcgacatcccaaatgcgcaa	4681
Db	4081	aaggccagtgatatcgctgtctcatctaaacaccttatctggcgacatcccaaatgcgcaa	4140
Qy	4682	ggggcaagccaagcaaaccaactcagcagctatgtgatctgatggcaataaaggtcatc	4741
Db	4141	ggggcaagccaagcgaacacactcagcagctatgtggatgtgtggtatggcaataaaggtcatc	4200
Qy	4742	tatgacagtaccgataacaagctactatcaagccaaaaatgatggccagcttgataaaacc	4801
Db	4201	tatgacagtaccgataacaagctactatcaagccaaaaatgatggccagcttgataaaacc	4260
Qy	4802	aaagaagttgccaaagacaaaactggtgcgccaaagcccaacccccagatggcaatgtggct	4861
Db	4261	aaagaagttgccaaagacaaaactggtgcgccaaagcccaacccccagatggcaatgtggct	4320
Qy	4862	caaatgaatgtcaaatcagctcatcaacaagaacaagtaaatgatgccataaaaagcaa	4921
Db	4321	caaatgaatgtcaaatcagctcatcaacaagaacaagtaaatgatgccataaaaagcaa	4380
Qy	4922	ggcatcaatgaagacaaacgctctgttaaaggacttgaaaagcgctctctgataacaaa	4981
Db	4381	ggcatcaatgaagacaaacgctctgttaaaggacttgaaaagcgctctctgataacaaa	4440
Qy	4982	accaaaaacgcgcgagtaactgtgggtgatttaaatgcgcttgcccaaacacgcgtgacc	5041
Db	4441	accaaaaacgcgcgagtaactgtgggtgatttaaatgcgcttgcccaaacacgcgtgacc	4500
Qy	5042	tttgacggggatcacggcacaacgcgtaaaaaactggcgagagactttgaccatacaaggct	5101
Db	4501	tttgacggggatcacggcacaacgcgtaaaaaactggcgagagactttgaccatacaaggct	4560
Qy	5102	gggcaaacagacaccaataagctaacccgataataacatcggtgtggtagcaggtactgat	5161
Db	4561	gggcaaacagacaccaataagctaacccgataataacatcggtgtggtagcaggtactgat	4620
Qy	5162	ggcttcaactgtcaaaactgtgccaaagacctaaccatacttaacagcgttaatgcaggtggc	5221
Db	4621	ggcttcaactgtcaaaactgtgccaaagacctaaccatacttaacagcgttaatgcaggtggc	4680
Qy	5222	accaaaatgatgacaaaaggctgtctttgtagactcaacgcgtcaagcccaagaagcaaac	5281
Db	4681	accaaaatgatgacaaaaggctgtctttgtagactcaacgcgtcaagcccaagaagcaaac	4740
Qy	5282	acccctgtctaaagtgcctaattggcttgacctgggtggcgaaggttcatcagtaattgtgggc	5341
Db	4741	acccctgtctaaagtgcctaattggcttgacctgggtggcgaaggttcatcagtaattgtgggc	4800
Qy	5342	aaaggcacaanaagataccgacgctgccattgtacacacagtttaacgaagctacgcaacttg	5401
Db	4801	aaaggcacaanaagataccgacgctgccattgtacacacagtttaacgaagctacgcaacttg	4860
Qy	5402	ttgggtcttgtaagtctggtaataataacgctgacggcaatcagggtaaacattgcgcac	5461
Db	4861	ttgggtcttgtaagtctggtaataataacgctgacggcaatcagggtaaacattgcgcac	4920
Qy	5462	atcaaaaaagaccaaaatctcaggttcatcatctaaccgactgttcaacaagcaggcacg	5521
Db	4921	atcaaaaaagaccaaaatctcaggttcatcatctaaccgactgttcaacaagcaggcacg	4980
Qy	5522	gtacttggcgttaaaagtaataacgataccgaaaaaacttgcacatggctgggtatatacaagt	5581

Db	4981	 gtacttgcggttaaagtaataaacagatccggaataacttgcactggttggtatacaagtg
Qy	5582	ggcgtgataaaacacgcgaacgctaacgcgatttaaagcaatgtttgggtcaaaacccaa
Db	5041	ggcgtgataaaagcagcgcaacgctaacgcgatttaagcaatgtttcgggtcaaaacccaa
Qy	5642	aaagatggcagcaaaaaagccctgtcgcgaactataacgcgcaggtcgagaccaactat
Db	5101	aaagatggcagcaaaaaagccctgtcgcgaactataacgcgcaggtcgagaccaactat
Qy	5702	ttgacaaacaaacccgcagaagccattgacagaaataatgaacaaagttatccgtttcttc
Db	5161	ttgacaaacaaacccgcagaagccattgacagaaataatgaacaaagttatccgtttcttc
Qy	5762	catgtcaacgatggcaatcaagagccctgtggtacaagggcgtaacgcgattgactcaagt
Db	5221	catgtcaacgatggcaatcaagagccctgtggtacaagggcgtaacgcgattgactcaagt
Qy	5822	gcctcagggcaagcactcagtgggcgataggtttccagggccaaaggcgagatggtagagcgcc
Db	5281	gcctcagggcaagcactcagtgggcgataggtttccagggccaaaggcgagatggtagagcgcc
Qy	5882	gttgccataggcgagcaaaacccaaagcagggaacccaatccatcgccatcgggtgataacgca
Db	5341	gttgccataggcgagcaaaacccaaagcagggaacccaatccatcgccatcgggtgataacgca
Qy	5942	caagccacggcgatcaatccatcgccatcggtacaggaagaatgtgtagcaggttaagcac
Db	5401	caagccacggcgatcaatccatcgccatcggtacaggaagaatgtgtagcaggttaagcac
Qy	6002	tctggtgccatcggcgacccaagcactgttaaggtgataaacagttacagttggtggttaat
Db	5461	tctggtgccatcggcgacccaagcactgttaaggtgataaacagttacagttggtggttaat
Qy	6062	aaacaccagtttacgatgccactaaacccgatgtcttgggtggggcaataaacatcacc
Db	5521	aaacaccagtttacgatgccactaaacccgatgtcttgggtggggcaataaacatcacc
Qy	6122	gtgacgcaagaagtaactcgggttcgttaggttcaaaactctgccatcagtgcaaggcacacac
Db	5581	gtgacgcaagaagtaactcgggttcgttaggttcaaaactctgccatcagtgcaaggcacacac
Qy	6182	gcagggcacacaagcccaaaaaatctgacgcgacagcaggtacaacaccacacagcaggtgca
Db	5641	gcagggcacacaagcccaaaaaatctgacgcgacagcaggtacaacaccacacagcaggtgca
Qy	6242	accggtacggttaaaggcttctgtgacaaaacggcggttgggtgcggttccggtgggtgcc
Db	5701	accggtacggttaaaggcttctgtgacaaaacggcggttgggtgcggttccggtgggtgcc
Qy	6302	tcaagttctaaagccgataccaaaatgtggcagcaggtgaggtcagtgccaccagcacc
Db	5761	tcaagttctaaagccgataccaaaatgtggcagcaggtgaggtcagtgccaccagcacc
Qy	6362	gatcggttcaatggttagccagttgtacaagccaccaaaagcattgccaaacgcacaacaaat
Db	5821	gatcggttcaatggttagccagttgtacaagccaccaaaagcattgccaaacgcacaacaaat
Qy	6422	gagcttgaccatcgtataccaccaaaacgaaaaataaggccaatgcaagggtatttcacgag
Db	5881	gagcttgaccatcgtataccaccaaaacgaaaaataaggccaatgcaagggtatttcacgag
Qy	6482	atggcgatggcggtccatgccaagaacctacattctctggcagatccataggttaccgggggt
Db	5941	atggcgatggcggtccatgccaagaacctacattctctggcagatccataggttaccgggggt
Qy	6542	attgccaccacaacggtcaaggtgcggtggcagtgaggactgcgaagctgcgaagctgcgaataat
Db	6001	attgccaccacaacggtcaaggtgcggtggcagtgaggactgcgaagctgcgaagctgcgaataat
Qy	6602	ggttcaattgggtattttaaaatacaatggttcagccgatccccaaggccatgtagggggcgca

Db 6061 ggtcaatgggtatttaaaatcaatggttcaagccgatacccaaggccatgtagggcgga 6120

Qy 6662 gttggtgcaggttttcaacttt 6682

Db 6121 gttggtgcaggttttcaacttt 6141

RESULT 15

US-08-621-944-2

; Sequence 2, Application US/08621944

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Ken

; APPLICANT: HARKNESS, Robin E.

; APPLICANT: LOOSMORE, Sheena M.

; APPLICANT: KLEIN, Michael H.

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER

; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/621,944

; FILING DATE: 26-MAR-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/478,370

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-587

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5976 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-621-944-2

Query Match 85.7%; Score 5976; DB 10; Length 5976;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 707 gtgatcggtgcaacgctcagtgccagtgcttctgctcaaaaaaagatacccaacatc 766

Db 1 GTGATCGGTGCAACGCTCAGTGCGGCGTGTATGCTCAAAAAAAGATACCAACATATC 60

Qy 767 gcaattggtgaacaaaccagccagcgtcagggcctgcaaggcgagcggtgatcga 826

Db 61 GCAATTGCTGTAACAAACCAAGCAAGCGCTCAGGCACTGCCAAGCGCGAGCGGTGATCGA 120

Qy 827 gcaattggtgtaaaatgctaaacgacagcggtcagacccatcgccatcggtgag 886

Db 121 GCCATGCTATTGTTGAAATGCTTAACGCGCAGCGCGGTCAAGCCATCGCCATCGGTAGT 180

Qy 887 agtaataaaactgtcaatggagcagtttgatagaataggtaccggtgctacgggtcaa 946

Db 181 AGTAATAAAACTGTCAATGGAAGCAGTTTGGATAAGATAGTACCGATGCTACGGGTCAA 240

Qy 947 gagtccatcgccatcggtggtgataaaaggctagtggtgctcctcgattgcccaggt 1006

Db 241 GAGTCCATCGCCATCGGTGATGTAAGGCTAGTGTGATGCTCGATTGCCATCGGT 300

Qy 1007 agtgcacttacatcttgcctgatacagcagtggttaactcctcaaacatccgaaggtactcg 1066

Db 301 AGTGATGACTTACATTTGCTTGATCAGCATGGTAATCCTAAACATCCGAAGGTACTCTG 360

Qy 1067 attaacgatcttataacggccatcgagttataaagaatacgaagctcaaaaggataat 1126

Db 361 ATTAAACGATCTTATAACGGCCATGCGAGTATTAAGAATAACAGCTCAAGGATAAT 420

Qy 1127 gatgtaaaatatacagccacacacccgcaagcgagccacccagtcagtgagggccatg 1186

Db 421 GATCTAAAATATAGACGCACAAACCGCAAGCGGACACGCCAGTACTGCAGTGGGAGCCATG 480

Qy 1187 tcatagcacagggctatttttccaaagcctttggtacacggcgacacagctcaaaagtcgc 1246

Db 481 TCATATGCACAGGGTCAATTTTCCAAACGCTTTGGTACACGGGCAACAGCTTAAAGTGCC 540

Qy 1247 tattcctggcagtggtcttgcgcgcacacagcgagggcccaatctacaatcgctattggt 1306

Db 541 TATTCTTGGCAGTGGGTCTTGGCGGCACACGCGGAGGCCAATCTACAATCGCTATTGGT 600

Qy 1307 tctgatgcaacatctagctcgttggggagcgatagcccttggcgaggtactcgtgctcag 1366

Db 601 TCTGATGCAACATCTAGCTCGTTGGAGCGATAGCCCTTGGTGCAGGTACTCTGCTCAG 660

Qy 1367 ctacagggcagttatgcccctaggttcaaggttctgtgtcactcagagtgataataattct 1426

Db 661 CTACAGGGCAGTATTGGCCCTAGGTCAAGTTCTGTTGCTCCTCAGAGTGATTAATAATTCT 720

Qy 1427 agaccggcctatacaccaaaataccagggcactagaccaccaagtttcaagccaccaataat 1486

Db 721 AGACCGGCTTATACACCAAAATACCCAGGCAGTACAGCCCAAGTTTCAAGCCACCAATAAT 780

Qy 1487 acgaaggcggttccactttccattggttagtaactctatcaaacgttaaatcatcaatc 1546

Db 781 ACGAAGCGGGTCCACTTTCCATTGCTAGTAACTCTATCAAAACGTAAATCATCATGTC 840

Qy 1547 ggtgcaggtgttataaaacccgagtcggtcgaatgtggcacagctagaagcggtggtgaag 1606

Db 841 GGTGCAGGTGTTAATAAAACCGATCGGTCAATGTGGCACAGCTAGAAAGCGGTGTGAAG 900

Qy 1607 tgggctaaggcgctagaaataacttttcaggggtgatacagctacgactgacgtgaaaaata 1666

Db 901 TGGGCTAAGGAGCGTAGAATTTACTTTTCAGGGTGATGATAACAGTACTGACGTAAATA 960

Qy 1667 ggttggataataacttacttaactataaagggtggtgcagagaccaacgcattaccgataat 1726

Db 961 GGTGTTGGATAATACTTTAACTTAAAGGTGCTGCAGAGACCAACGCATTAAACCGATAAT 1020

Qy 1727 aatacgggtgtgtaaaaggcgctgataatagtggtctgaaagtttaacttgcctaaact 1786

Db 1021 AATATCGGTGTGTAAGAGAGGCTGATAATAGTGTGCTGTAAGTTTAACTTCTGCTAAAAC 1080

Qy 1787 ttaacaaacttactgaggtgatacaactacataataatgcccacacacagcttaaggtca 1846

Db 1081 TTAACAACTTCTACTGAGGTGATGATACACTATTAAATGCCCAACACCCACAGTTAAGGTA 1140

Qy 1847 ggtagtagtagtagtactacagctgaattattgagtgatagtttaacctttaccagccc 1906

Db 1141 GGTAGTAGTAGTAGTACTACAGCTGAATTTATTGAGTGATAGTTTAACTTTTACCAGGCC 1200

Qy 1907 aatacaggcgagtcgaaggcaagcaaacccgctctatggcgttaagtgggtgagtttact 1966

Db 1201 AATACAGGAGCTCAAGAGCAAAACCGCTATGGCGTTAATGGGTGAGGTTTACT 1260

Qy 1967 aataatgcagaacacacagcagcaatcggcactactcgtattaccagagataaaattggc 2026

Db 1261 AATAATGCAGAAACACACAGCAGCAATCGGCAGTACTCTGTTATTACCAGAGATAAAAATTGSC 1320

QY 2027 ttgtctgagatggtgtagtggatgaagaaacaaagcaccatatttggataaaaaaacatt 2086
DB 1321 TTGTGCTGAGATGGTGTGATGATAAACAAGCACCATATTTGGATAAACAACATT 1380
QY 2087 aaagtgggtgtagtggcaattaccatagacaaatgcatgtgtagcaggtaataaaaaagac 2146
DB 1381 AAAGTGGGTAGTGTGCCAATTACCATAACAAATGSCATTTGATGCAAGGTAAATAAAGATC 1440
QY 2147 agtaattcttgccaaaggtagcagtgctaacgatgcggtttaccatgaaacagctcaaaagcc 2206
DB 1441 AGTAATCTTGCCAAAGGTAGCAGTGCTAACGATGCGGTTACCATCGAACAGCTCAAGGCC 1500
QY 2207 gccaaagctactttaaacgagggcgctggatcagtgatcaacctactgaaatatcagtt 2266
DB 1501 GCCAAGCCTACTTTAAACGCGAGCGCTGGCATCAGTGTACACACCTTACTGAAATATACATT 1560
QY 2267 gatcgaagagtggaattgttacgcgcacacacttacacattggcgtgaaacacacagag 2326
DB 1561 GATGCTAAGAGTGGCAATGTTACCGCCCACTTACACATTTGGCGTGAAACACACCGAG 1620
QY 2327 cttaacagtgatggcactagtgataaaatttagtggtaagggttagtgggtacgaacaatagc 2386
DB 1621 CTAAACAGTGATGGCACTAGTATATAATTTAGTGTAAAGGTAGTGGTACGAACAATAGC 1680
QY 2387 ttagtaccgcgcgaacatttggcaagctatctaaatgaagtaacatcgaaacggctgacagt 2446
DB 1681 TTAGTTACCGCGGAACATTTGGCAAGCTATCTAAATGAAGTCAATCGAACGGCTGACAGT 1740
QY 2447 gctctacaaagcttaccgtttaaagaagaagcagatgatgcgcgaacgcctatcacctg 2506
DB 1741 GCTCTACAAAGCTTTACCGTTAAAGAAAGACGATGATGACCGCAACGCTATCACCGTG 1800
QY 2507 gctaaagatacgaacaaaaatgcccgcgcagtcagcatctttaaactcaaaagggtaaaaac 2566
DB 1801 GCTAAAGATACGACAAAAATGCCGGCGCAGTCAGCATCTTAAACTCAAAGGTAAAAAC 1860
QY 2567 ggtctaaagggtgtctacaaaaaaagatggtacggttacctttgggcttagccaagatagc 2626
DB 1861 GGTCTAACGGTGTGTACCAAAAAGATGGTACGGTTACCTTTGGGCTTAGCCAAAGATAGC 1920
QY 2627 ggtctgaccattggcaaacgaccctaaacacagatggtgactgttaaagataccaac 2686
DB 1921 GGTCTGACCATTTGGCAAAAGCACCTTAACACAGATGCTTGACHTTAAAGATACCAAC 1980
QY 2687 gaacaaatccaaagtgcgttaagtggcattaaattactaatgtgaatggtagtaacca 2746
DB 1981 GAACAAATCCAAAGTCGGTGCTAATGGCATTAATTTTACTAATGTGAATGGTAGTAATCCA 2040
QY 2747 ggtactggcattgcaaatcacgcctgcgaattaccagagataaaattggctttgctgtct 2806
DB 2041 GGTACTGGCATTTGCAAAATACCGCTCGCATTACCAGAGATAAAATTTGGCTTTGCTGGTTCT 2100
QY 2807 gatgctgcagttgatcaacaaacacttatcttgatcaagcaagctacaagttggcaat 2866
DB 2101 GATGCTGCAGTTGATACAAACAAACCTTATCTTGTATCAAGACAACTACAAGTTGGCAAT 2160
QY 2867 gttaaagattaccaacactggcattaacgcaggtggttaaagccatcacaggcgtgtcccca 2926
DB 2161 GTTAAAGATTACCAACACTGGCATTAACGCAGGTGTTAAAGCCATCACAGGGGTGTCCCCA 2220
QY 2927 acactgcctagattgcgcgatacaaaagtgcgcgcaacatagaaactgggcaatacaatccaa 2986
DB 2221 ACACCTGCCTAGATTGGCCGATCAAAAGTAGCCGCAACATAGAACTGGGCAATACAAATCCA 2280
QY 2987 gacaagacaaatccaacgcctgccagcatatgatataataataacagcctttaacccta 3046
DB 2281 GACAAGACAAATCCAAACGCTGCCAGCATTAATGATATATTAATATACAGGCTTTAACCTA 2340
QY 3047 aaaaaataaacaacccattgactttgtctccacttatgacattgttgacttttggccaat 3106
DB 2341 AAAAAATAAACAACCCATTGACTTTGTCTCCACTTATGACATTTGTGACTTTGCCAAT 2400
QY 3107 ggcgaatgccaccacgcacagtaaccatgataccgctaaacaaacacagtaaaagtggta 3166

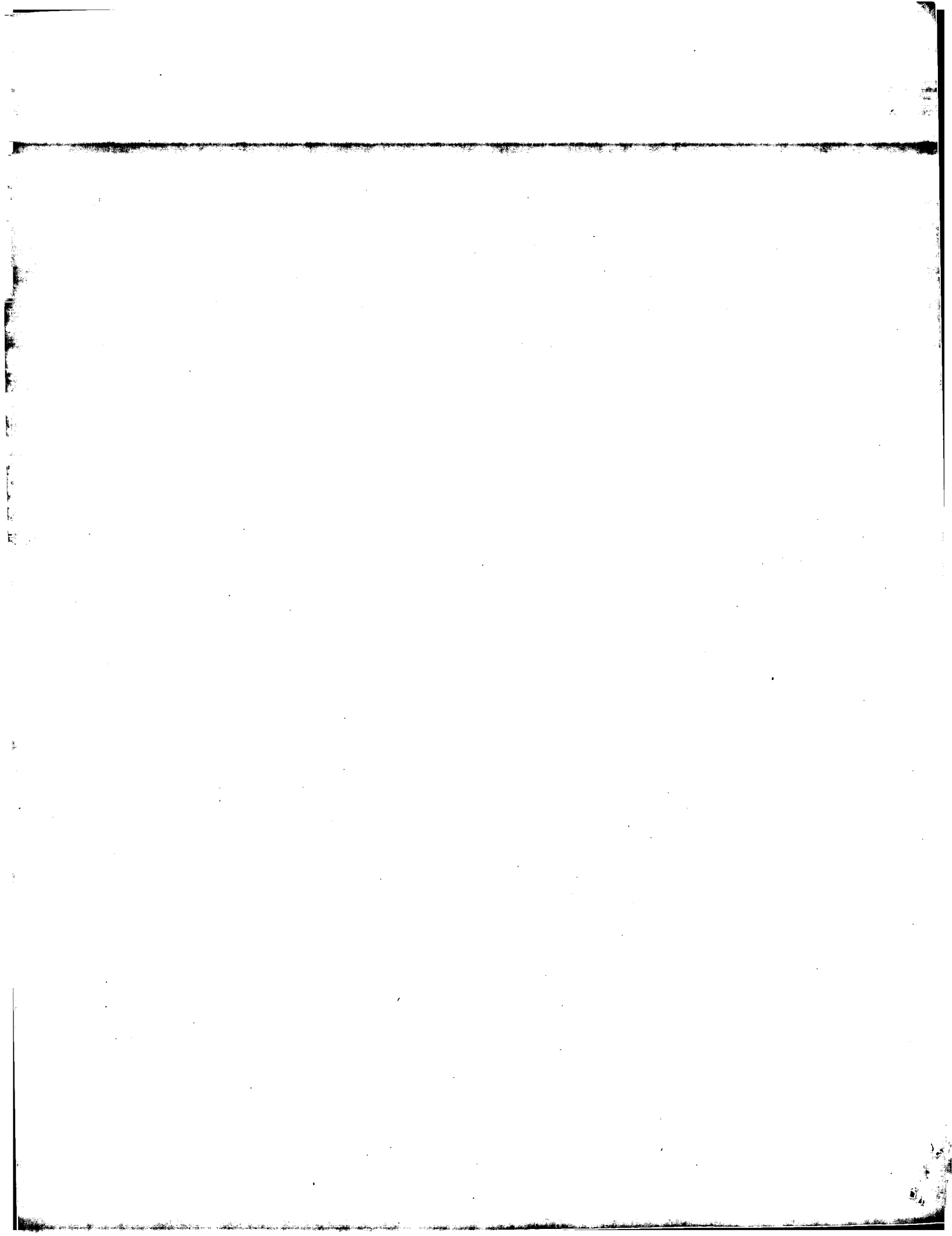
DB 2401 GGCAATGCCACCGCCACAGTAACCCATGATACCGTTAAACAAAACAGTAAGTGGTA 2460
QY 3167 tatgtgtgaatgtgatgatacaaccattcatctaaacagcactgatgacaataaaaaa 3226
DB 2461 TATGATGTGAATGTGATGATACACCACTTTCATCTAACAGGCACATGATGACAAATAA 2520
QY 3227 ctggcggtcaaaacaccacaaactgaacaaacaagtgttaattggttaatacagcaactaac 3286
DB 2521 CTTCGCGTCAAAACCCACCAACTGAACAAAACAAGTGTATATGTTAAATACAGCAACTAAC 2580
QY 3287 tttaagttaactctagttagaagatgcccttgttaacgccaagacatcgccgaaat 3346
DB 2581 TTTAATGTAACTCTAGTGTGAAGTGCCTTGTTAACGCCAAAAGACATCGCCGAAAT 2640
QY 3347 cttaaaccccttagccaagaaattcacaccacaaaggcacagcagacacccgcctacaa 3406
DB 2641 CTAAACACCTTAGCCAAAGAAATTCACACCACCAAGGACAGCAGACACCGCCCTACAA 2700
QY 3407 accttaccgttaaaaggtagatgaaataataatgtctgatgacgccaacgccaatcacc 3466
DB 2701 ACCTTTACCGTTAAAAGGTAGATGAAAATAATAATGTGTGATGACGCCAACCCATCAC 2760
QY 3467 gtgggtcaaaagaaacgcaaaataatcaagtcaaacacccctaaactcaaaagggtgaaacggt 3526
DB 2761 GTGGGTCAAAAGAACGCAAAATAATCAAGTCAACACCTTAACACTCAAAGGTGAAAACGCT 2820
QY 3527 cttaataataaaacccgacaaaatgtgacggtttacctttggcattaaacacacagcgt 3586
DB 2821 CTTAATATTAACCCGACAAAATGTGACGGTTACTCTTTGGCATTAACACCAACACGGT 2880
QY 3587 cttaagccggcaaaagcacccctaaacacggtggtctgtctattaaaaaccccactggt 3646
DB 2881 CTTAAAGCCGGCAAAAGCACCTTAACGACGCTGGCTGTCTATTAAAAACCCCACTGCT 2940
QY 3647 agcgaaacaaatccaaagtcggtgctgatggcgtgaagtttgcgaagtttaataatggt 3706
DB 2941 AGCGAACAAATCCAAAGTCGGTGCTGATGGCGTGAAGTTTGCCCAAGTTTAATAATAGT 3000
QY 3707 gttgtaggtgctgctgattgtagcacaactcgctattaccagagatgaattggttact 3766
DB 3001 GTTGTAGGTGCTGGCATTTGATGGCACAACTCGCATTTACCAGAGATGAATTTGGCTTTACT 3060
QY 3767 gggactaatggctcacttgataaaagcaaacccacccctaaagcaaaagcggcattacgca 3826
DB 3061 GGGACTAATGGCTCACTTGATAAAAGCAAAACCCCACTTAAGCAAAAGACGGCATTAAGCA 3120
QY 3827 ggtgtaaaagattaccaacattcaatcaggtgagattggcccaaaacagccatgct 3886
DB 3121 GGTGTAAAAGATTACCAACATTCATCAGGTGAGATTGCCCAAAACAGCCATGATGCT 3180
QY 3887 gtgacagcgcgcaagatttatgattaaaaacccgaacttgaacaaacaaatcgagctact 3946
DB 3181 GTGACAGCGCGCAAGATTATGATTTAAAACCCGAACCTTGAACAAACAAATTCAGCAGTACT 3240
QY 3947 gccaaaaacagcaaaaactcattacacgaattctcagtagcagatgaacaaggttaataac 4006
DB 3241 GCCAAAACAGCACAAAACCTCATTACAGAAATCTCAGTAGCAGATGAACAAGGTAAATAAC 3300
QY 4007 tttaagggttagtaaccccttactccagttatgacaccccaagacccctctgactatcacc 4066
DB 3301 TTTACGGTTAGTAACCCCTTACTCCAGTTATGACACTCAAGACCTCTGATGTCATCAC 3360
QY 4067 ttgtaggtgaaaaacgcatcaccaccccaaggtlaataaagggtggtgctggtgagcatt 4126
DB 3361 TTTGACAGGTGAAAACCGCATTTACCAACCAAGGTAAATAAAGGTGTGGTGGCTGTGGCATT 3420
QY 4127 gacaaacaaagggttaaccacgcctgaactgacgctgggtggtggttaataatgcaaaagc 4186
DB 3421 GACCAACCAAAAGGCTTAACCAACGCTTAAGCTGACCGTGGGTAAATAATATGCAAGGC 3480
QY 4187 attgtcattgacagcgaataatggtcaaaataaccatacagaggactgaacaaacttagct 4246

Db 3481 ATTCTCATTTGACACGCCAAATGGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCT 3540
 Qy 4247 aatgttaccaatgataaaggtagcgtaacgacacacagaggaaggaatataataataaagac 4306
 Db 3541 AATGTTACCAATGATTAAGGTAGCTAGCGACACACAGACAGGCAATATAATCAAGAC 3600
 Qy 4307 gaagacaaacccgctgcccagcagcatgttgatgtgctgaagcgagcgttttaacttgcaa 4366
 Db 3601 GAAGACAAAACCCGTCGCCGACGATTTGATGTCTAAGCGCAGGCTTTAACTTGCA 3660
 Qy 4367 gcaatgtgaacggttaacttctccacttatgacacgctcaactttgccgatgac 4426
 Db 3661 GGCAATGGTGAAGCGGTGACTTTGCTCCACTTTATGACACCGTCAACTTTGCCGATGSC 3720
 Qy 4427 aatgccaccacccgtaagggtgacctatgatgacacagcaaacaccagtaaaagtgtctat 4486
 Db 3721 AATGCCACCCGCTAAGGTGACCTATGATGACACAGCAAGCAAAACCACTAAGGTGCTAT 3780
 Qy 4487 gatgtcaatgtgatgatacaaccattgaagttaagataaaaaacttggcgtataaacc 4546
 Db 3781 GATGTCAATGTGGATGATACACCACTTGAAGTTAAGATTAANAACCTTGGCGTAAANACC 3840
 Qy 4547 accattgaccagctactggcagagtgctaaataaattggccctaaagcaatcaagctact 4606
 Db 3841 ACCACATTTGACCACTACTGGCACAGGTGCTTAATAAATTTGCCCTAAGCAATCAAGCTACT 3900
 Qy 4607 ggcgatgcgcttgtaaggccagtgatcgttgcctcaataaacacaccttatctgacgac 4666
 Db 3901 GGCGATGCGCTGTGCAAGCGCAGTGATGCTGCTCATCTAATCAACACCTTTATCTGGCGAC 3960
 Qy 4667 atccaaactgcgaaggcgcaagcgaagcgaacaaactcagcagcgtatgtggatgctgat 4726
 Db 3961 ATCCAACTGCCAAGGGGCAAGCAAGCGCAACCAACTCAGCAGGCTATGTGGATGCTGAT 4020
 Qy 4727 ggaataaggctcatctatgacagctaccgatacaagctactatcaagcacaataatgatgac 4786
 Db 4021 GGCAATAAGGTCACTATGACAGTACCGATTAACAAGTACTATCAAGGCAAAATATGTGGC 4080
 Qy 4787 acagttgataaaacaaagagttgcaaaagacaaactggcgcgaagcccaaacccca 4846
 Db 4081 ACAGTTGATAAACCNAAGAGTTGCCAAGACAAACTGGTGCCTCAAGGCCCAAAACCCCA 4140
 Qy 4847 gatggacattggctcaaatgaatgtcaaatcagtcattcaacaaagcaacagtaaatgat 4906
 Db 4141 GATGGCACATTTGGCTCAAAATGAATGCTAAATCAGTCAATTAACAAGAACAAATGAATGAT 4200
 Qy 4907 gcaataaaacaaagcagcatcaatgaagcaaacgccttgttaaaggacttgaaaaagcc 4966
 Db 4201 GGCAATAAAGCAAGGCATCAATCAAGCAACACGCTTTGTTAAAGGACTTGAAGAGCC 4260
 Qy 4967 gcttctgataacaaacaaacccgagcagtaactgttggttgatttaaatgcccgttgc 5026
 Db 4261 GCTTCTGATAAACAACCAACCAACGCGCAGTAACCTGTGGGTGATTTAAATGCGGTTGCC 4320
 Qy 5027 caaacccgctgacctttgcaggggatacagcacaacgcttaaaaaactggcgagact 5086
 Db 4321 CAAAACCCGCTGACCTTTGCAGGGGATACAGGCACAACGGCTTAANAACCTGGCGAGACT 4380
 Qy 5087 ttgacatcaaaaggtggcaacagacaccaaataagctaacccgataataacatoggtgtg 5146
 Db 4381 TTGACCATCAAGGTGGCAACACAGACCAATAAGCTAAGCTAATAACATCGGTGTG 4440
 Qy 5147 gtagcagggtactgatggttctactgtcaaaacttgcaaaagacctaaccaacttaacagc 5206
 Db 4441 GTAGCAGGTACTGATGGCTTCACTGTCAAACTTGGCAAGACCTTAACCAATTTAACAGC 4500
 Qy 5207 gtaatgcaggtggcaccacaaatgatgacaaaggcgtgtctttttgtagactcaagcgtg 5266
 Db 4501 GTTAATGCAGGTGGCACCACAAATTTGATGACAAAGGCGTGTCTTTGTAGACTCAAGCGGT 4560
 Qy 5267 caagccaaagcaaacacccctgtgctaaagtgcataatggcgtggacctgggtgcaaggtc 5326
 Db 4561 CAAGCCAAAGCAACACCCCTGTGCTAAGTGGCAATGGGCTGGACCTGGGTGGCAAGGTCT 4620

Qy 5327 atcagtaattgtgggaaaggcacaaaaagataccgaagcgtgccaatgtacaacagtttaaac 5386
 Db 4621 ATCAGTAAATGTGGCAAAAGGCACAAAAGATACCGACGCTGCCAATGTACAAAGTTAAAC 4680
 Qy 5387 gaagtcagcaacttgttgggtcttggtaatactgtggtaataataacgctgacgcaatcag 5446
 Db 4681 GAAGTCAGCAACTTTGTTGGGTCTTTGGTAAATGCTGGTAAATGATACGGCTGACGCAATCAG 4740
 Qy 5447 gtaaacattgcccagacatacaaaaagaccaaaatccaggttcatcatcctaaccgcaactgac 5506
 Db 4741 GTAAACATTTGCCGACATCAAAAAAGACCCCAATTCAGGTTTCATCATCTAACCGCACGTGC 4800
 Qy 5507 atcaaaagcagcagcgtacttggcggttaaggtaataaacgataccgaaaaacttgcacact 5566
 Db 4801 ATCAAAAGCAGCAGCGTACTTTGGCGGTAAAGGTAAATACGATACCGAANAACCTTGGCACT 4860
 Qy 5567 ggtggtatacaagtggtggttgataaaagacgcaacgctaacgctgatttaagcaatggtt 5626
 Db 4861 GGTGGTATACAAAGTGGCGGTGGATAAGACCGCAACGGTAACGGCGATTTAAAGCAATGTT 4920
 Qy 5627 tgggtcaaaaccccaaaaagatggcagcaaaaagccctgctgcgcaactataacgcgcga 5686
 Db 4921 TGGGTCAAAACCCCAAAAAGATGGCAGCAAAAAGCCCTGCTCGCACATTAATACGCCGCA 4980
 Qy 5687 ggtcagacaaactatttgacacaaaccccgacagaagcattgacagaataaataagcaaa 5746
 Db 4981 GGTCAAGCAACTATTTTGACCAACAAACCCCGCAGAACGCTTTGACGAGTAATTAATGACNA 5040
 Qy 5747 ggtatccgcttctccatgtcaacgattggcaatcaagcgctgtggtacaaggcgtaac 5806
 Db 5041 GGTATCCGCTTCTTCCATGTCAACCATGGCAATCAAGACGCTGTGTTACAAAGGGGTAAAC 5100
 Qy 5807 ggcattgactcaagtgccctcaggaagcactcagtggcgataagtttccagggcaagca 5866
 Db 5101 GGCATTGACTCAAGTGCCTTCAGGCAAGCACTCAGTGGCGATAGGTTTCAGGCCCAAGCA 5160
 Qy 5867 gatgtgaagccgcgctgacctagcagacaaaacccagcaggaacaaacccatccatcgc 5926
 Db 5161 GATGGTGAAGCCGCTTGCCTATAGCAGACAAACCCAAAGCAGGCAACCAATCCATCGCC 5220
 Qy 5927 atcgggtataagcacaagccagcggtatccatccatccatccatccatccatccatccatcc 5986
 Db 5221 ATCGGTGATAAACCAAGCCACGGGCGATCAATCCATCCCATCGGTACAGGCAATGTG 5280
 Qy 5987 gtagcaggttaagcactctgtgacctcgcgcacccaagcactgttaagcgtgataaact 6046
 Db 5281 GTAGCAGGTAAAGCAGCTCTGTGTGCCATCGGGACCCCAAGCAGCTGTAAAGCTGATAACAGT 5340
 Qy 6047 tacagtggtggttaataacacccggtttacgatagcactcaaaacgactgtctttggtgtg 6106
 Db 5341 TACAGTGTGGGTAAATAACACCACTTTACGATGCGCACTCAAAACCGATGCTTTGGTGTG 5400
 Qy 6107 ggaataacataccgtgacctgacctgacctgacctgacctgacctgacctgacctgacctg 6166
 Db 5401 GGCAATAACATCAACCTGACCGGAAAGTAACCTCGTTGCGCTTAGTTCAAACTCTGCCATC 5460
 Qy 6167 agtgaggcacaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6226
 Db 5461 AGTGCAGGCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 5520
 Qy 6227 accacagcaggtgcaacccggttaaggttgaaggttgcctgacacaaacgagcagcagcagc 6286
 Db 5521 ACCACAGCAGGTGCAACCGGTACGGTTAAAGGCTTTGCTGGCAACACGGCGGTGGTGGC 5580
 Qy 6287 gtcctcgtggtgctcaggtgctgaacgagcagcagcagcagcagcagcagcagcagcagc 6346
 Db 5581 GTCTCCGTGGGTGCTCAGGTGCTGAACGCGGTATCCAAATGTGGCAGCAGGTGAGGTC 5640
 Qy 6347 agtgcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6406
 Db 5641 AGTGCCACCAAGCAGCGGTGCGGTCAATGGTGAAGCAGGTGTACAAAGCCACCAAGCATT 5700

Qy	6407	gccaacgcaaccaatgagcttgaccatcgatatccaccacaaacgaaaaataagggccaatgca	6466
Db	5701	gccaacgcaaccaatgagcttgaccatcgatatccaccacaaacgaaaaataagggccaatgca	5760
Qy	6467	gggatttcatcagcgatggcgatggctccatgcccacaaagcctacattcctggcgagatcc	6526
Db	5761	gggatttcatcagcgatggcgatggctccatgcccacaaagcctacattcctggcgagatcc	5820
Qy	6527	atggttaccgggggtattgcccaccacacacggtcaaggtgcggtggcgagtgggactgtcg	6586
Db	5821	atggttaccgggggtattgcccaccacacacggtcaaggtgcggtggcgagtgggactgtcg	5880
Qy	6587	aagctgtcgggataatggtcaatgggtatttaaaatcaatggttcagccgatacccaaggc	6646
Db	5881	aagctgtcgggataatggtcaatgggtatttaaaatcaatggttcagccgatacccaaggc	5940
Qy	6647	catgtaggggcgcgatggttggtgcaggttttcaacttt	6682
Db	5941	catgtaggggcgcgatggttggtgcaggttttcaacttt	5976

Search completed: September 13, 2001, 11:47:58
Job time: 83299 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 13:00:38 ; Search time 3832.77 Seconds
(without alignments)
3470.029 Million cell updates/sec

Title: US-09-361-619-5
Perfect score: 6972
Sequence: 1 catggatagggcagggtgt.....ccgtgacctgcaggtcgac 6972

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1639839 seqs, 953802495 residues

Total number of hits satisfying chosen parameters: 3279678

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents.NA.New.*

- 1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.2	0.9	1797	5	US-09-771-382-18
2	58.6	0.8	1797	5	US-09-771-382-19
3	58.2	0.8	1815	5	US-09-771-382-22
4	57	0.8	1224	5	US-09-771-382-30
5	57	0.8	1302	5	US-09-771-382-31
6	57	0.8	1509	5	US-09-771-382-32
7	57	0.8	1539	5	US-09-771-382-28
8	57	0.8	1770	5	US-09-771-382-14
9	57	0.8	1776	5	US-09-771-382-12
10	57	0.8	1776	5	US-09-771-382-16
11	57	0.8	1779	5	US-09-771-382-21
12	55.4	0.8	1785	5	US-09-771-382-15
13	55.4	0.8	1785	5	US-09-771-382-20
14	55.4	0.8	1800	5	US-09-771-382-17
15	53.8	0.8	1542	5	US-09-771-382-29
16	53.8	0.8	1779	5	US-09-771-382-13
17	49.6	0.7	889	5	US-09-838-601-2719
18	46.4	0.7	96046	5	US-09-803-736-274
19	46.2	0.7	481	5	US-09-688-848-5933
20	44.8	0.6	547	5	US-09-850-147-8017
21	44.2	0.6	666	5	US-09-838-601-2717
22	44.2	0.6	834	5	US-09-823-241-8800
23	44	0.6	70559	5	US-09-409-800B-1
24	44	0.6	114144	5	US-09-803-736-325
25	43.8	0.6	1083	1	PCT-US01-08631-12549
26	43.8	0.6	1146	5	US-09-803-110-2020
27	43.6	0.6	58985	5	US-09-901-152-3

28	43.6	0.6	73391	5	US-09-803-736-1313
29	42.8	0.6	480	5	US-09-821-710-4611
30	42.6	0.6	1131	5	US-09-803-110-2030
31	42.6	0.6	3145	5	US-09-788-657-9
32	42.6	0.6	1408	5	US-09-788-657-12
33	42.4	0.6	1250	5	US-09-803-110-1859
34	42.4	0.6	100887	5	US-09-803-736-1269
35	42.2	0.6	2163	5	US-09-760-475-1749
36	42.2	0.6	20420	5	US-09-764-905-27977
37	42.2	0.6	107865	5	US-09-803-736-1375
38	42	0.6	1879	5	US-09-765-272-105
39	41.8	0.6	713	5	US-09-838-601-2714
40	41.8	0.6	918	5	US-09-588-018-3
41	41.8	0.6	89473	5	US-09-803-736-994
42	41.6	0.6	90170	5	US-09-803-736-1321
43	41.4	0.6	293	5	US-09-813-155-1599
44	41.4	0.6	580	5	US-09-850-147-9636
45	41.4	0.6	1770	5	US-09-906-514-1

ALIGNMENTS

RESULT 1

US-09-771-382-18
; Sequence 18, Application US/09771382
; GENERAL INFORMATION:
; APPLICANT: Peak, Ian
; APPLICANT: Jennings, Michael
; TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
; FILE REFERENCE: 8795-2401
; CURRENT APPLICATION NUMBER: US/09/771.382
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 60/177,917
; PRIOR FILING DATE: 2000-01-25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-09-771-382-18

Query Match	0.9%	Score 60.2;	DB 5;	Length 1797;
Best Local Similarity	52.1%	Pred. No. 9.3e-05;		
Matches 134;	Conservative 0;	Mismatches 123;	Indels 0;	Gaps 0;
Qy 6429	accatcgatccacacaaacgaaataaggccaaatgcaggagatttcacagcgatggcga	6488		
Db 1541	acaacggcatcgacaatgtgacgcaacgcgcgcggtatcgcccaagcgattgcaa	1600		
Qy 6489	tggcgatccatccacacgctacattcctgcagatccatggtaccgggggtattgcca	6548		
Db 1601	ccgcagggttgccacagcgctattgcccggcaagatgatgagcgatcgccgcgata	1660		
Qy 6549	cccacacgggtcaagggtcggtggcagtgggacgtgcgagctgcgataatgtcaat	6608		
Db 1661	cttatcgcgggcgaagcgggttacgccatcgctactcagcatttctgacactgggaatt	1720		
Qy 6609	gggtatttaaatcaatggttcacgccgataccccaggccatgtagggcgccagttgtg	6668		
Db 1721	gggtattcaaggcagcgcttcggcaattcgcgccattcggttgcgttcgcgcatcg	1780		
Qy 6669	caggttttcaacttttaa	6685		
Db 1781	tcggttatcatggtgtaa	1797		

RESULT 2

US-09-771-382-19
; Sequence 19, Application US/09771382
; GENERAL INFORMATION:

[illegible]

	Query Match	0.88;	Score 57;	DB 5;	Length 1539;
	Best Local Similarity	51.4%;	Pred. No.	0.00056;	
	Matches 132;	Conservative 0;	Mismatches 125;	Indels 0;	Gaps 0;
Qy	6429	accatcgtatccacaaacgaaaataaggccaatgcaggagatttcacagcgatggcga	6488		
Db	1283	acaaccgatcgcaaatgtgcagcgcaacgcgctgcggcatgcccaagcgattgcaa	1342		
Qy	6489	tggcgctccatgccaaagcctacattcctgcagatccatggttacgggggtattgccca	6548		
Db	1343	ccgcaggctcgttcaggcgctatttgcgccgcaagagtatgatgcgatcgcgccgca	1402		
Qy	6549	cccacaacggttcaagggtgcggtgcagtgggacctgtcgaaagctgtcgagataatggtcaat	6608		
Db	1403	cttatcgcgcggaagcggttcacgccatcgctactccagtatttcgcacgcgcggaatt	1462		
Qy	6609	gggtattttaaatcaatggttcagccgcgatacccgaagcccatgtaggggcgcgagttggtg	6668		
Db	1463	ggattatcaaggcaagcgcttcgggcaattcgcggccattcgggtgcttccgcactgt	1522		
Qy	6669	caggttttcaacttttaa	6685		
Db	1523	tcgattatcaagtataa	1539		

```

RESULT      8
US-09-771-382-14
; Sequence 14, Application US/09771382
; GENERAL INFORMATION:
; APPLICANT: Peak, Ian
; APPLICANT: Jennings, Michael
; TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
; FILE REFERENCE: 8795-24U1
; CURRENT APPLICATION NUMBER: US/09/771,382
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 60/177,917
; PRIOR FILING DATE: 2000-01-25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 1770

```


Query Match	0.88;	Score 57;	DB 5;	Length 1776;
Best Local Similarity	51.4%;	Pred. No. 0.0006;		
Matches 132;	Conservative	0;	Mismatches 125;	Indels
			0;	Gaps
Qy	6429	accatcgtatccaccacaacgaaataaggccaatgcagggaattcatcagcgaatggcga	6488	
Db	1520	acaaacgcgatcgacaattggacggcaacgcgcgtggcgatcgcccaagcgattggcaa	1579	
Qy	6489	tgcgctccatgccaacgcctacattcctctgcagatccatggtttaccgggggtattgccaa	6548	
Db	1580	cgcgaggctcgtttcaggcgattttgcgccggcaagagtatgatgcgatcgcggcgcgcaa	1639	
Qy	6549	cccaacaacgttcaagttcgggtggcgatgggacatgtcgaaactgttcggataatggtccaat	6608	
Db	1640	cttatcggggcgaagccgggttacgccatcggctactccagtatttcgcacggcggaaatt	1699	
Qy	6609	gggtattttaaataatcgaattggttcagccgataccaagcccatgtaggggcggcagttgggtg	6668	
Db	1700	ggattatcaaggcagcggttcctccgaattcgcggcaattcgcggcatttcgggtgcttcgcgactcg	1759	
Qy	6669	caggttttccacttttaa	6685	

RESULT 11
 US-09-771-382-21
 Sequence 21 Application US/09771382
 GENERAL INFORMATION:
 APPLICANT: Peak, Ian
 APPLICANT: Jennings, Michael
 TITLE OF INVENTION: MODIFIED SURFACE ANTIGEN
 FILE REFERENCE: 8795-2401
 CURRENT APPLICATION NUMBER: US/09/771,382
 PRIOR FILING DATE: 2001-01-25
 CURRENT APPLICATION NUMBER: US 60/771,917
 PRIOR FILING DATE: 2000-01-25
 NUMBER OF SEQ ID NOS: 52
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 21
 LENGTH: 1779
 TYPE: DNA
 ORGANISM: Neisseria meningitidis
 US-09-771-382-21

Query Match	0.8%;	Score 57;	DB 5;	Length 1779;
Best Local Similarity	51.4%;	Pred. No. 0.0006;		
Matches 132; Conservative		0;	Mismatches 125;	Indels 0; Gaps 0;

Search completed: September 13, 2001, 12:35:56
Job time: 84918 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2001, 13:49:24 ; Search time 78.73 Seconds
(without alignments)
1576.239 Million cell updates/sec

Title: US-09-361-619-7
Perfect score: 10303
Sequence: 1 MNHYKVFNRATGTFMAVA.....NGSADTQCHVCAAVGAGPHF 2047

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10303	100.0	2047	22 AAB69134	M. catarrhalis str
2	10024	97.3	1992	22 AAB69133	M. catarrhalis str
3	10021	97.3	1992	17 AAW04505	Moraxella 200 kDa
4	10021	97.3	1992	22 AAB69137	M. catarrhalis M56
5	7675	74.5	2053	22 AAB69135	M. catarrhalis str
6	6593.5	64.0	2314	22 AAB69136	M. catarrhalis les
7	1123	10.9	2353	17 AAR99393	Haemophilus influe
8	1109	10.8	2411	21 AAB23860	Haemophilus influe
9	1016	9.9	1104	21 AAB23856	Haemophilus influe
10	1016	9.9	1104	21 AAB23859	Haemophilus influe
11	970	9.4	1004	21 AAB23857	Haemophilus influe

12	927.5	9.0	1002	21 AAB23854	Haemophilus influe
13	813	7.9	2042	19 AAW56319	Haemophilus paraga
14	804	7.8	2039	19 AAW56322	Haemophilus influe
15	697.5	6.8	1094	21 AAB23858	Haemophilus influe
16	678.5	6.6	1098	17 AAR99392	Haemophilus adhesi
17	573	5.6	679	17 AAR99394	Haemophilus adhesi
18	573	5.6	679	21 AAB23855	Haemophilus influe
19	520.5	5.1	1601	18 AAW30292	Non-typeable Haemo
20	505.5	4.9	1598	18 AAW30291	Non-typeable Haemo
21	496.5	4.8	1529	14 AAR41732	High molecular wei
22	478.5	4.6	2383	21 AAB15945	E. coli proliferat
23	477.5	4.6	2514	21 AAV75097	Neisseria meningit
24	462	4.5	2893	19 AAW98828	H. pylori GPHO 148
25	462	4.5	2893	19 AAW71556	Helicobacter polyp
26	460.5	4.5	1978	20 AAV72720	Amino acid sequenc
27	457	4.4	1338	14 AAR41731	High molecular wei
28	452	4.4	2599	21 AAV75098	Neisseria meningit
29	451	4.4	3647	11 AAR05041	Filamentous haemag
30	450.5	4.4	1638	20 AAY00138	Enterococcus faeca
31	450.5	4.4	1638	20 AAY00140	Enterococcus faeca
32	450.5	4.4	1638	20 AAY00142	Enterococcus faeca
33	448	4.3	3596	21 AAY87407	Bordetella pertuss
34	443.5	4.3	1962	12 AAR10563	Mutant protease (K
35	443	4.3	1221	21 AAB01825	Haemophilus influe
36	443	4.3	1227	21 AAB01824	Haemophilus influe
37	439.5	4.3	1962	12 AAR10560	Mutant protease (K
38	439.5	4.3	1968	12 AAR10941	Mutant protease (d
39	438.5	4.3	1962	12 AAR10561	Mutant protease (N
40	438.5	4.3	1981	19 AAW42634	Protein sequence t
41	438	4.3	1959	12 AAR10562	Mutant protease (d
42	437.5	4.2	1962	12 AAR10557	Mutant protease (A
43	436.5	4.2	1962	12 AAR10558	Mutant protease (A
44	436.5	4.2	1962	12 AAR10559	Mutant protease (A
45	435	4.2	1612	19 AAW65088	R. prowazekii S-la

ALIGNMENTS

RESULT 1

AAB69134

ID AAB69134 standard; Protein; 2047 AA.

XX AAB69134;

XX 24-APR-2001 (first entry)

XX M. catarrhalis strain 4223 genomic 200kda protein SEQ ID NO:7.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;
XX 200kda outer membrane protein; antibacterial; immunogenic; infection;
XX otitis media; detection.

OS Moraxella catarrhalis.

PN WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59102, AAF59103.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis

Claim 1; Fig 3A-W; 247pp; English.

PS The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (AB), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis strain 4223 genomic 200kDa protein, which is given in the CC exemplification of the present invention.

XX Sequence 2047 AA;

Query Match 100.0%; Score 10303; DB 22; Length 2047;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNHLYKVFENKATGTFMAVEAKSHSTGGSCATQGVSVCTLSFARLAALVLYIGAT 60
DB 1 MNHLYKVFENKATGTFMAVEAKSHSTGGSCATQGVSVCTLSFARLAALVLYIGAT 60
QY 61 LSGSAYAQKDKTHIAIGBQNPRRSGTAKADGADRAIAIGENAGGQAIAGSSNKTV 120
DB 61 LSGSAYAQKDKTHIAIGBQNPRRSGTAKADGADRAIAIGENAGGQAIAGSSNKTV 120
QY 121 NGSSLDKIGTDATQBSIAIGDVKASGDASTAIGSDDLHLDDQHGNPKHPKGTTLNDLI 180
DB 121 NGSSLDKIGTDATQBSIAIGDVKASGDASTAIGSDDLHLDDQHGNPKHPKGTTLNDLI 180
QY 181 NGHAVLKEIRSSKNDNVKYRRYRTASGHASTAVCAMSYAOGHFSNAPGTRATAKSAYSLAV 240
DB 181 NGHAVLKEIRSSKNDNVKYRRYRTASGHASTAVCAMSYAOGHFSNAPGTRATAKSAYSLAV 240
QY 241 GLAATAEGOSTIAIGSDATSSSLGATLACAGTRAQIQQSGSIALCGSVVTVQSDNNRSPAYT 300
DB 241 GLAATAEGOSTIAIGSDATSSSLGATLACAGTRAQIQQSGSIALCGSVVTVQSDNNRSPAYT 300
QY 301 PNTQALDPKFOATNPKAGPLSTGSNSIKRKIINVAGYKNTKDANVAQLEAVVWAKER 360
DB 301 PNTQALDPKFOATNPKAGPLSTGSNSIKRKIINVAGYKNTKDANVAQLEAVVWAKER 360
QY 361 RITFGDDNSTDVKIGLDNLTIKGGAETNALTDNNIGVVKADNSGLKVKLAKTLNNLT 420
DB 361 RITFGDDNSTDVKIGLDNLTIKGGAETNALTDNNIGVVKADNSGLKVKLAKTLNNLT 420
QY 421 EVNTTTLNATTVKVGSSSTTAELLSDSLSTPTQNTGSGQSTKTVYGVNGYKFTNNAFT 480
DB 421 EVNTTTLNATTVKVGSSSTTAELLSDSLSTPTQNTGSGQSTKTVYGVNGYKFTNNAFT 480
QY 481 TAAIGTRITTRDKIGFARDGDVDEKQAPYLDKQKLVKGSVAITIDNGIDAGNKKISNLAK 540
DB 481 TAAIGTRITTRDKIGFARDGDVDEKQAPYLDKQKLVKGSVAITIDNGIDAGNKKISNLAK 540
QY 541 GSSANDAVTIEQLKAAPPTLNAGAGTSVPTPTETISVDKSGNVTAPTNYIGVKTTELSNSG 600
DB 541 GSSANDAVTIEQLKAAPPTLNAGAGTSVPTPTETISVDKSGNVTAPTNYIGVKTTELSNSG 600
QY 601 TSDKFSVKGSGTNNLSVTAELIASYLNENVRTPADSALQSFTVKEEDDDANAITVAKDPT 660
DB 601 TSDKFSVKGSGTNNLSVTAELIASYLNENVRTPADSALQSFTVKEEDDDANAITVAKDPT 660
QY 661 KNAGAVSILKLGKNGLTVAATKDKGTVTGFLSQDSGLRTIGKSTLNDGLTKDTNEQIQV 720
DB 661 KNAGAVSILKLGKNGLTVAATKDKGTVTGFLSQDSGLRTIGKSTLNDGLTKDTNEQIQV 720

QY 721 GANGIKFTVWNGSNPOTGTANTARITRDKIGFAGSDGAVDTNKNPYLDQDKLQVGNVKITN 780
DB 721 GANGIKFTVWNGSNPOTGTANTARITRDKIGFAGSDGAVDTNKNPYLDQDKLQVGNVKITN 780
QY 781 TGINAGSKAITGLSPPLPSIADQSSRNIEGNTIQQDKSNAASINDILNTGFENLKNNN 840
DB 781 TGINAGSKAITGLSPPLPSIADQSSRNIEGNTIQQDKSNAASINDILNTGFENLKNNN 840
QY 841 PIDFVSTYDIVDFANGNATTATVTHDTANKTSKVVDVNVDDTTIHLTGTDDNKKLGVT 900
DB 841 PIDFVSTYDIVDFANGNATTATVTHDTANKTSKVVDVNVDDTTIHLTGTDDNKKLGVT 900
QY 901 TKLNKTSANGNTATNPNVNSDEDALVNAKDAENLTLAKEIHTPKGPADTALQTFVYK 960
DB 901 TKLNKTSANGNTATNPNVNSDEDALVNAKDAENLTLAKEIHTPKGPADTALQTFVYK 960
QY 961 KYDENNNADANAITVGOKNANNQVNTLTKGENGLINIKTDKNGTVTFGINTTSLGKAGK 1020
DB 961 KYDENNNADANAITVGOKNANNQVNTLTKGENGLINIKTDKNGTVTFGINTTSLGKAGK 1020
QY 1021 STLNDGGLSIKNPTGSEIQVGADGVYKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTNGS 1080
DB 1021 STLNDGGLSIKNPTGSEIQVGADGVYKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTNGS 1080
QY 1081 LDKSPHLSKDGINAGGKKITNIQSGEIAQNSHDVATGCKIYDLKTELENKISSAKTAQ 1140
DB 1081 LDKSPHLSKDGINAGGKKITNIQSGEIAQNSHDVATGCKIYDLKTELENKISSAKTAQ 1140
QY 1141 NSLHFSFVADEGGNFTVSNPYSSYDTSKTSKVITFAGENGITTKVNKGVRVGDQTKG 1200
DB 1141 NSLHFSFVADEGGNFTVSNPYSSYDTSKTSKVITFAGENGITTKVNKGVRVGDQTKG 1200
QY 1201 LTPKLTGVGNNKGIVIDSQNGQNTITGLSNTLANVTNNDKGSVRTTEOGNIKDEKTR 1260
DB 1201 LTPKLTGVGNNKGIVIDSQNGQNTITGLSNTLANVTNNDKGSVRTTEOGNIKDEKTR 1260
QY 1261 AASIVDVLISAGFNLOGCEAVDFVSTYDTPNFADGNATTAKTVYDDTSKTVKVVVDVND 1320
DB 1261 AASIVDVLISAGFNLOGCEAVDFVSTYDTPNFADGNATTAKTVYDDTSKTVKVVVDVND 1320
QY 1321 DTTIEVKDKLGKVTTLTSTGTGANKFALSNOATGDLAKASDIVAHLNLTSGDITQAK 1380
DB 1321 DTTIEVKDKLGKVTTLTSTGTGANKFALSNOATGDLAKASDIVAHLNLTSGDITQAK 1380
QY 1381 GASQANNAGYVDADGKNVYDSTDNKYQAQNDGTVDKTKEVAKDKLVAQAQTPDGLA 1440
DB 1381 GASQANNAGYVDADGKNVYDSTDNKYQAQNDGTVDKTKEVAKDKLVAQAQTPDGLA 1440
QY 1441 OMNVKSVLNKEQVNDANKKOGINEDNAPVKGLEKAASDNKTKNAATVYGDNLNAVAQTPLT 1500
DB 1441 OMNVKSVLNKEQVNDANKKOGINEDNAPVKGLEKAASDNKTKNAATVYGDNLNAVAQTPLT 1500
QY 1501 FAGDGTAKLGETLTIKGGQTDNTKLTDDNNIGVYVAGTGTGFTVKLAKDLTLNLSNVNAGG 1560
DB 1501 FAGDGTAKLGETLTIKGGQTDNTKLTDDNNIGVYVAGTGTGFTVKLAKDLTLNLSNVNAGG 1560
QY 1561 TKIDDKGVSVFVSDSSQAKANTPVLISANGLDGCKVIVSNVKGKTKDTDAANVOQLNEVRNL 1620
DB 1561 TKIDDKGVSVFVSDSSQAKANTPVLISANGLDGCKVIVSNVKGKTKDTDAANVOQLNEVRNL 1620
QY 1621 LGLGNAGNDNADGNOVNITADIKDPNSGSSSNRTVYKAGTVLGGKGNNDTEKLTATGGIQV 1680
DB 1621 LGLGNAGNDNADGNOVNITADIKDPNSGSSSNRTVYKAGTVLGGKGNNDTEKLTATGGIQV 1680
QY 1681 GVDKGNANGDLSNVVVKTKDGSKKALLATYNAAGQNTYLTNNPAAEIDRINEQIRFF 1740
DB 1681 GVDKGNANGDLSNVVVKTKDGSKKALLATYNAAGQNTYLTNNPAAEIDRINEQIRFF 1740
QY 1741 HVNDGNQPVVQGRNGIDSSASGKHSVAIGFOAKADGAAVAIGRQTQAGNQSIAIGNA 1800
DB 1741 HVNDGNQPVVQGRNGIDSSASGKHSVAIGFOAKADGAAVAIGRQTQAGNQSIAIGNA 1800
QY 1801 QATGDSIAIGTVNVVACKHSGAIGDPSTVKADNSYSVGNNNQFTDATQTDVDFGVGNIT 1860

Db 1801 qatdqsiaigtgnvqgkhsaigdbstvkadnsysvgnnmqfdatqtdvfgvgnmit 1860
 QY 1861 VTESNSVALGSNSAISAGTHAGTQAKSDSTAGTGTAGATGTVKGFAGTAVGAVSVGA 1920
 Db 1861 vtesnsvalgsnsaisagthagtqakksdgtagtttttagatgtvkgfaggtavgavsvga 1920
 QY 1921 SGAERRTONVAGVSVSATSTDAVNGSOLYKATQSIANATNELDHRHQNENKANAGISSA 1980
 Db 1921 sgaerrtonvavgevsatstadvngsglykatgsianatneldhrhqnknagiss 1980
 QY 1981 MAMASMPQAYIPGRSMVTGGIATHNGOGAVAGLSKLSNDGOWYFKINGSADTQGHVGA 2040
 Db 1981 mamasmppayipgrsmvtggiathngogavagvlgsklsndgqvwfkingssadtqghvga 2040
 QY 2041 VGAGFHF 2047
 Db 2041 vgagfhf 2047

RESULT 2

AAB69133
 ID AAB69133 standard; Protein; 1992 AA.
 XX AC AAB69133;
 XX DT 24-APR-2001 (first entry)
 XX DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.
 KW Moraxella catarrhalis strain 4223; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; detection.
 XX OS Moraxella catarrhalis.
 XX PN W0200107619-A1.
 XX PD 01-FEB-2001.
 XX PF 26-JUL-2000; 2000WO-CA00870.
 XX PR 27-JUL-1999; 99US-0361619.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 DR WP: 2001-159722/16
 DR N-PSDB: AAF59100, AAF59101.
 XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis
 XX Example 3; Fig 2A-W; 247pp; English.

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is used in the exemplification of the present invention.

Sequence 1992 AA;

Query Match 97.3%; Score 10024; DB 22; Length 1992;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 VIGATLSSAYAOKKDKKHIAIGEQNQRSGTAKADGDRAIAIGENANAGGQAIATGS 115
 Db 1 vigaTlsgsayaqkdkthiaIgeqngqrrsgtakadgdraIaigenanaggqaiags 60
 QY 116 SNKTVNSSLKIGTDATQESIAIGGDVKGASGDASIAIGSDDLHLHQHNPXPKPTL 175
 Db 61 snktvngsslkigtatqesiaiggdvkgasgdasIaigsdldhlhqhnpkpkgtl 120
 QY 176 INDLINGHAVIKEIRSSKNDVKYRRRTASGHASTAVGAMSAQGHFSNAFCTRATAKSA 235
 Db 121 indlinghavIkeirsskndvkyrrrtasghastavgamsyaqghfsnafctratak 180
 QY 236 YSLAVGLAATAEGOSTIAIGSDATSSSLGATGALGAGTGAQILOGSIALQGSSVVTQSDNNS 295
 Db 181 yslavglAataegdstiaigsdatssslgaIalgaIagtraIqgssIalqgssvvtqsdnns 240
 QY 296 RPAYTPNTQALDPKFOATNNTKAGPLSIGSNSIKRKIINVAGVKNKTDVAVNVAQLEAVVK 355
 Db 241 rpaytpnqaldpkfoatnntkagplsignsikrkiinvagvknktdavnvavleavvk 300
 QY 356 WAKERRITFOGDDNSTDVKIGLDNTLTIKGAETNALTDNNIGVYKEADNSGLKVKLAKT 415
 Db 301 wakerriTfOgddnstdvkiGLDntltikgaetnaltdnnigvYkeadnsglkvklakt 360
 QY 416 LNNITEVNTTLNATTVKVGSSSSSTAEELSDSLTFTQPNVTGSGSTSKTVYGVNGVKFT 475
 Db 361 lnnitevnttlNatTVkvgsssstaeelSdslftqpnvtgsgstsktvYgvngvkft 420
 QY 476 NNAETTAAGTTRTRDRKIGFARGDGDVDEKQAPYLDKKQKLVGSVAITIDNGIDAGNKKI 535
 Db 421 nnaettaagTtrtrdrkiGfargdgdvdeKqapYldkkqkLvgsvaltidngidagnk 480
 QY 536 SNLAKGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISVDKSGNVTAPTNTYNGVKTE 595
 Db 481 snlkgssandavtIeqLkaakptlnagagisvtpTeisvdKsgnvtapntYngvkTte 540
 QY 596 LNSDGTSDKFSVKGSGTNNSLVTAHSLASYLEVNRPTADSLQSFVKEEDDDANAITV 655
 Db 541 lnsdgtSdKfsvkgsgTnnsLvtaehLasylevnrPtadslqSfVkeeedddanaItv 600
 QY 656 AKDTTKNAGAVSILKLGKNGLTVATKKDGTVTFGLSODSGLTIGKSTLNDGLTVKDTN 715
 Db 601 akdttknagavSIlkLgkngltvatKkdgtvtfGlsodsgltigKstlnndglTvkDtn 660
 QY 716 BQIQVGANGIKFTNVNGSNPGTGIANTARITRDKITGFAGSDGAVDTNKPYLDDKQLQVGN 775
 Db 661 eqiqvgangikftnvngsnpgtgiAntaritrDKitgfagSdgavdtNkpyldqkqlvg 720
 QY 776 VKINTGINAGGKAITGLSPILPSIADQSSRNIELGNTIQDKDSNAASINDILNTGNL 835
 Db 721 vkiIntginaggkaiTglspilPSiAdqssrnieLgntIQdkdsnaasindilntgnl 780
 QY 836 KNNNNPIDFVSTYDIDVPFANGNATVTHDTANKTSKVYDVNVVDVDTIHLTGTDNNKK 895
 Db 781 knnnpidfVstydIdvpfAngnatvthDtAnktskvYdvNVdVdtihlTgtdnnk 840
 QY 896 LGVKTTLINKTSANGTATNFNVNSDEDAVLVNARDIAENLTAKEIHTTKGTADTALQ 955
 Db 841 lgvkttlinktsangtatnfnvnsdEdaVlVnArDiAenltaKEihttkgtadtalq 900
 QY 956 TFFTVKVDENNADANAITVQKNNANQVNTLTLKGENGLNIKTDKNGTVTFGINTTSG 1015
 Db 901 tftvkvdennadanaItvqKnnanqVntltlKgenGLniKtdKngtvTFginttsg 960
 QY 1016 LKACKSTLNDGGLSLKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDCTTRTRDEIGFT 1075
 Db 961 lkagkStlndgglsLknptgSeqiqvgadgvKfakvNngvvgagIdcttrtrdeIgf 1020

QY 1076 GTNGSLDKSPHLSKOGINAGGKKTINIOGGEIAQNSHDAVTCGKIYDKTELENKISST 1135
 Db 1021 gtngsldkskphlskdglnagggklniqgselaqnsdhavtggkiydktelenkisst 1080
 QY 1136 AKTAQNSLHEFSVADRGNNFTVSNPNYSYSDTSKTDVTFAGENGITTKVKNKGVVVRVGI 1195
 Db 1081 aktqnsllhefsvadeggnftvsnpyssydsksktdvtfagengitckvknkgvvrvgi 1140
 QY 1196 DQTKGLTTPKLTVGNNGKGIIVDSQNGQNTITGLSNTLANVTNDRSGSVRTTEQGNIIKD 1255
 Db 1141 dqtkgltpkltvgnngkgividsngqntitglsntlanvtnndksgrvtteqgnlikd 1200
 QY 1256 EDKTRAASIVDLVSAGFNLOGNEAVDFSVTYDTPVNPADGNATTAKVTYDDTSKTSKVY 1315
 Db 1201 edktraasivdlvsagfnlqngneavdfsvtydtpvnpadgnattakvtyddtsktskvvy 1260
 QY 1316 DVNVDDTTEVDDKLGKVTTLTSTGTGANKPALNSQATGDAVKASDVAHLNLSGD 1375
 Db 1261 dvnvddttievddkklgvttltstgtgankfalnsqatgdalvksadivahlntlsdg 1320
 QY 1376 IOTAKASQANNSAGYVDADGNKVIYDSDNKYIQAKNDGTVDKTKEVAKDKLVAQAQTP 1435
 Db 1321 iotakasqannsagvydadgnkvlydsdnkyyqakndgtvdktkevdkklvagaqtp 1380
 QY 1436 DGTLAQNNVKSVINKEQVNDANKKQGINEDNAPVKGLEAASDNKTKNAAVTVGDLNAVA 1495
 Db 1381 dgtlaqnnvksvinkeqvndankkqginednapvkglekaasdnktknaavtvgdlnava 1440
 QY 1496 QTPLTTFAGDTGTTAKLIGFTLTIKGGOTDTNKLTDNNIGVAVGCTDGTVKLAKDLNLNS 1555
 Db 1441 qtplttfagdtgttackligeltikggotdtnkltdnnigvavgtgftvklakdlnlns 1500
 QY 1556 VNAGGKTIDDKGVSYFVDSGQKANTPVLSANGLDLGGKVISNVGKGTDKTDAANYQQLN 1615
 Db 1501 vnaggktiddkgvsvfvsdgqkancpvlsangldlgkvisnvvgkgtktdaanyqqln 1560
 QY 1616 EVRNLLGLGNAGNDADGNQVNIADIKDPNSGSSNRVIRAGTVLGGKGNNDTEKLTAT 1675
 Db 1561 evrnllglnagnadnqgnvniadlkkdpnsgssnrvtvikagtvlgkgnndteklat 1620
 QY 1676 GGTQGVGDKNANGDLSNVWKTQKDGSKALLATYNAAAGOTNYLTNNPAEALDRINQ 1735
 Db 1621 ggtqgvgdknangdlsnvwktqdkgskallatynaagotnyltnpaealdrineq 1680
 QY 1736 GIRFFHVNDGNQBPVYVGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRTOAGNQSTA 1795
 Db 1681 girffhvnngnqbpvvyvgrngidssasgkhsvaigfqakadgeaavaigrqtqagnqsta 1740
 QY 1796 IGDNAQATGDOSTAICTGNVAVGKHSGATCDPSTVKADNSYSVGNNGQTTDATQTDVFGV 1855
 Db 1741 igdnaqatgdstaictgnvavgkhsgatcdpstvkadnsysvgnngqttdatqtdvfgv 1800
 QY 1856 GNNITVTESNVALGNSAITSAGTHAGTQAKKSDGTAGTTTATAGTGVKGFAGQTPAVGA 1915
 Db 1801 gnnitvtesnvalgnsaitsagthagtkaksksdgtagttttagtcgvkfgagqtpavga 1860
 QY 1916 VSVGASGAERRIQNVAAGEYSATSTDAVNGSQLYKATQSTANATNLDHRIHQENKANA 1975
 Db 1861 vsvgasgaerriqnvaageysatstdavngsqlykatqstianatnldhrihqenkana 1920
 QY 1976 GISSAMAMASMPQAYTPGRSMVTGGTATHNGQAVAVGLSKLSDNGQWFKINGSDTQOG 2035
 Db 1921 gissamamasmqaytpgrsmvtggathngqavavglsklsgdnqgwfkingsadtqg 1980
 QY 2036 HVGAAYVAGGFHF 2047
 Db 1981 hvgaayvagghf 1992

RESULT 3
 AAW04505

ID AAW04505 standard; Protein; 1992 AA.

XX AAW04505;
 XX AC
 XX DT 25-JAN-1997 (first entry)
 XX DE Moraxella 200 kDa outer membrane protein.
 XX KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;
 KW diagnosis.
 XX OS Mycobacterium catarrhalis strain 4223.
 XX PN W09634960-A1.
 XX PD 07-NOV-1996.
 XX PF 29-APR-1996; 96WO-CA00264.
 XX PR 26-MAR-1996; 96US-0621944.
 PR 01-MAY-1995; 95US-0431718.
 PR 07-JUN-1995; 95US-0478370.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
 XX DR WPI; 1996-506162/50.
 DR N-PSDB; AAT38740.
 XX Moraxella outer membrane protein - useful as immunogen in protective
 PT vaccine and for diagnosis
 XX Claim 14; Fig 6; 109pp; English.
 XX An approx. 200 kDa outer membrane protein (AAW04505) can be
 CC isolated from Moraxella catarrhalis otitis media strain 4223
 CC by electroelution, or expressed from a gene (see also AAT38740)
 CC obtd. from a strain 4223 genomic library. Natural or recombinant
 CC outer membrane protein is useful as an immunogen to protect
 CC against infection by Moraxella, esp. M. catarrhalis. It can
 CC also be used to detect antibodies, esp. for differential diagnosis
 CC between bacteria that cause similar symptoms, and also useful as
 CC a carrier for other antigens and used to raise antitumour
 CC antibodies for conjugation to therapeutic agents.
 XX SQ Sequence 1992 AA;

Query Match 97.3%; Score 10021; DB 17; Length 1992;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1991; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 56 VTGATLSSAYAQKDKTHIAIEQNPRRSGTAKADGDRATIAIGENANAQGGQATIGS 115
 Db 1 mgatlisgsayaqkdkthiaieqnprrsgtakadgdratiaigenanaqggqatigsg 60
 QY 116 SNKTVNGSLDKIGTDTATGOSTAIGDVKASGDASIAIGSDDLHLLDQHGNPKPKGTL 175
 Db 61 snktvngsldkigtatgostaiigdvkasgdasialgsddlhlldqgnpkpkgtl 120
 QY 176 IINDLINGHVLKEIRSSKNDVKYRPTTASGHASTAVGAMSVAQGHFSNAFCGTRATAKSA 235
 Db 121 iindlinghavlkeirsskndvkrrttasghastavgamsvaqghfsnaifgtrataka 180
 QY 236 YSLAVGLAATAFGQSTIAIGSDATSSSLGAIALGAGTQAQLOGSIALGSGSVVTTQSDNNS 295
 Db 181 yslavglataegqstiaigsdatslgaialgagtraqlqgsialgsgsvvtqsdnns 240
 QY 296 RPAYPTNTQALDPKFOATNTTAKPLSISGNSSTIKRKIIINVAGVNTKTDVAVNAQLEAVYK 355
 Db 241 rpaytntqaldpkfgatnttkagplsigsnsikrkliinvagvntkdavnvaqlaevyk 300
 QY 356 WAKERRITFGDDNSVDVKIGLDNTLTIKGGAETNALTDNNIGVVKKEADNSGLKVKLAKT 415

Db 301 wakerrifggdndstvdigldntltikgaecnaidnngivvkeadnsglkvklakt 360
Qy 416 LNNTEVNTTNNATTIVKVGSSSTTAELLSLSLTFTQPTNGSQSTKTYGVNGVKFT 475
Db 361 lnnitevnttlnattvkvsgsssttaeillsdltftqptngsqstektvvgvngvkt 420
Qy 476 NNAETTAAGTTRITROKIGFARGDGVDEKOAPYLDKKQLKGVSAITIDNGIDAGNKKI 535
Db 421 nnaettaaigttritrdkigfargdgvdekqpyldkkqlkvgsvaltidngidagnkkl 480
Qy 536 SNLAKGSSANDAVTEOLKAAPTLNACAGISVTPTEISVDKASGNVTAPYNTIGVKTTE 595
Db 481 snlkgssandavteqlkaaptlnacagisvtppteisvdaksngvncaptynigvktte 540
Qy 596 LNSDGTSDKFSVKGSGTNSLSVTAELHASYLNEVNRITADSALQSFTVKEEDDDANAITV 655
Db 541 lnsdgtstkfsvkgsgttnslsvtaehiasylnevnrtadsalqsftvkeedddanaitv 600
Qy 656 AKDTTKNAGAVSILKLGKNGLTVAATKKGTVTFGLSODSGLTIGCKSTPLNNDGLTVKDTN 715
Db 601 akdttknagavsiilkkgkngltvatkkdgtvtfglsgdsgltigkstinndgltvkdt 660
Qy 716 EOIQVGANGIKFTVWNGSNPCTGANTARITRDKIGFAGSDGAVDTNKPYLDDOKLQVCN 775
Db 661 eqiqvgangikftvwnsgnpgtganitaritrdkigfagsdgvotnkpylldqdklqvcn 720
Qy 776 VKIYNTGINAGKAITGLSPLTLPSTADQSSRNIELGNTIQDKDSNAASINDIILNTGNL 835
Db 721 vkintginagkaitglspplsiaqssrniegntiqdkdsnaasindiilntgfnl 780
Qy 836 KNNNPIDFVSTYDIYDFPANGNATVATVHTDANKTSKVYDVNVDDFTIHLTGFDNKK 895
Db 781 knnnpidfvtstydiydfpangnatvatvhtdanktskvdyvnnvddftihltgtdnkk 840
Qy 896 LGVTKTLNKTSAACNTATNENNVSSDEALVNAKDAENLTAKELHTTKGTADTALQ 955
Db 841 lgvtktlknktsangntatnenvnssdedalvnakdaenltakehlttkgtadtaiq 900
Qy 956 TFTVKVDENNADANAITYGQKANNQVNTLTIKGSEIAGNSHDVATGGKIYDLKTELENKISST 1015
Db 901 tftvkvdennadanaityvgkannqvntltikgenglniktdkngvtfngintsg 960
Qy 1016 LKAKSTLNDGGLSTKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFT 1075
Db 961 lkakstlndgglstknptgseqiqvgadgvkfakvnngvvgagidgtrtrtrdeigft 1020
Qy 1076 CTNGSLDKSPHLSKDGINAGGKKTNIQSGEIAQNSHDVATGGKIYDLKTELENKISST 1135
Db 1021 gtngsldkskphlskdginaggkktniqsgelagqnsdhdvttgkkydkltelenkist 1080
Qy 1136 AKTAQNSLREFSVADEQGNFTVSNPYSYDTSKTSQVITFAGENGITTKVKNKGVVRYGI 1195
Db 1081 aktagnslhefsvadeqgnftvsnpysydtstsdvittfagengittkvknkgvrvygi 1140
Qy 1196 DOTKGLTTPKTIUVGNNGKGVIDISQNGQNTITGLSNLTANVTNDKGSVRTTEQGNIIKD 1255
Db 1141 dqtkglttptkitvgnngkgvidsqngqntitglslntlanvtdnkgsvrtteqgniiikd 1200
Qy 1256 EDKTRAASIVDVLNAGFNWQNGEAVDFVSTYDTVNFADGNATTAKVYDDTSKTSKVY 1315
Db 1201 edktraasivdvlisagfnlgngeavdfvstydtvtnfadgnattakvtyddtsktskvvy 1260
Qy 1316 DVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNOATGDALVKASDIIVAHNLTLISGD 1375
Db 1261 dvnvddttievdkkkgvkttlstgtgankfalsngatgdalvkasdiivahnlntlsgd 1320
Qy 1376 IOTAKGASQANNSAGYVDADGNKVYIDSTDNKYQAKNDGTVDKTEKAKDLVAQAQTP 1435
Db 1321 iqtakgasqannsagyvddgnkvlyidstdnkyqakndgtvdktevakdklvaqaqtp 1380
Qy 1436 DGTLAQMMVKSVINKEQVNDANKQGINEDNAFVKGLEKASDNKTKNAAVTVGDNLNAYA 1495
|||||

Db 1391 dgtlaqmmvksvinkeqvndankkqginednafvkgkleasdnktnkaavtvvgdlnava 1440
Qy 1496 QTPLTTFAGDGTGTAKKLGELTITIKGGQTDITNKLTDPNNIGVVAGTDGFTVKLAKDLTNLNS 1555
Db 1441 qtpltfagdtgtakklgetitlikggqtditnkltannlgvvagtdgftvklakdltnlns 1500
Qy 1556 VNAGGTKIDDKGVSPVSDSSGQAKANTPVLSSANGLDLGGKVISNVGKGTDKTDAAANQQOLN 1615
Db 1501 vnaggtkiddkgvsfvdssgqakantpvlssangldlgkvksnvvgkctktdaanvqqln 1560
Qy 1616 EVRNLLGFGNAGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGKGNNDTEKLTAT 1675
Db 1561 evrnllgfgnagndnagdnqvniaidikkdpnsgssnrtvikagtvlggkgnndtektlat 1620
Qy 1676 GGIOGVGDKDNANGDLNVNVVKTQKDGSKKALLATYNAAGQNTYLTNNPAEIDRINEQ 1735
Db 1621 ggiovgvdkdngnangdlnvvvktqkdgskkallatynaagqntyltnnpaeaidrineq 1680
Qy 1736 GIRFFHVNDGNOEPVQQRNGIDSSASGKHSVAIQFQAKADGEAAVAIGRTQAOQNQSTIA 1795
Db 1681 girffhvndgnoepvqqrngldssasgkhsvaliqfakadgeaavaigrqtcagnqsl 1740
Qy 1796 IGDNAQATGDDOSIATGTGNVAGHSGAIGDPSTVKADNSYSVGNNOFTDATOTDVRGV 1855
Db 1741 igdnaqatgddgsiaigtgnvvagkhsgaigdpstvkadnsysvgnngftdatqtdvfgv 1800
Qy 1856 GNNITVTESNSVALGSNSAISAGTHAGTQAKSKDGTAGTATTAGATGTVKGFAGOTAVGA 1915
Db 1801 gnnitvtesnsvalgsnsaisagthagtcakksdgtagtttagatglvkgfagotavga 1860
Qy 1916 VSVGASGAEERRIQNVAAGEVSATSDAVNGSOLYKATQSIANATNELDHRITHQENKANA 1975
Db 1861 vsvgasgaerriqnvaaagevsatstdavngsqlykatqsianatneldhrilhqenkana 1920
Qy 1976 GISSAWANASMPQAVIPGRSMVTGGIATHNGGAVAGLSKLSNDGQWVKINGSADTQG 2035
Db 1921 gissamamampqayipgrsmvtggiaathnggavagvlsklsndgqwwfkingssadtqg 1980
Qy 2036 HVGAAGVAGGFHF 2047
Db 1981 hvgaavgagfhf 1992
RESULT 4
AAB69137
ID AAB69137 standard; Protein; 1992 AA.
XX AAB69137;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.
XX
KW Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
PN WO200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
DR WPI; 2001-159722/16.
DR N-PSDB; AAF59106.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis
PS Claim 1; Fig 8A-V; 247pp; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis M56 200kDa protein in pKS348, which is given in the
CC exemplification of the present invention.
XX
XX Sequence 1992 AA;

Query Match 97.3%; Score 10021; DB 22; Length 1992;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1991; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 56 VIGATLGSAYAKQDKTKHTAIGEQNPQRSGTAKADGDAIAGENANAQGGQATAIGS 115
DB 1 migtatlgssayackdktkhtiaigeqnqprsgtakadgdraiaigenanaqggqalaigs 60
QY 116 SNTVNGSSLDKICTGATGQESTAIAGDVKASGDASTAIGSDDLHLLDQGNPKPKGTL 175
DB 61 snktvngssldkigtatgesiaigdvkaskdasiasigsdhllldqgnpkpkgtl 120
QY 176 INDLINGHAVLKEIRSKDNVYRRTTASGHASTAVGAMSYAQGHFSNAGFTRATAKSA 235
DB 121 indlinghavlakeirskdnvyrtrttasghastavgamsyagghfsnafgtratak 180
QY 236 YSLAVGLAANTAEOSTTAIGSDATSSLSGAIAGAGTGAOLQGSIALGOGSVVYTSQDNNS 295
DB 181 yslavglantaesttaigsdatslsgaialagagtraglqslalqgsvvtsqdnns 240
QY 296 RPAYTPNTQALDPKFOATNNTKAGPLSIGNSIKRKTIINYGAGVKNTPADVNAQLEAVVK 355
DB 241 rpaytpntqaldpkfoatnntkagplsigsnsikrktiinygagvknktadvnvaql 300
QY 356 WAKERRITFQDDNSTDVKIGLWTLTIKGAETNALTNNIGVVKADNSGLKVLAKT 415
DB 301 wakeritfqqddnstdvkiglwtltikgaetnaltndnigvkvadnsglkvklakt 360
QY 416 LNNLTENVNTTLNATTIVKVGSSSTTAELLSLSLTTPQNTSQSTSKTVYGVNGVKFT 475
DB 361 lnnltevnttlnattivkvgssttaelldslsttpqntsqstsktvygvngvkft 420
QY 476 NNAETTAAGTRITRDRKIGFARDGVDEKQAPYLDKKQLKVGSVAITINGIDAGNKKI 535
DB 421 nnaettaagtritrdrkigfardgvdekoapyldkkqlkvgsvaltingidagnkki 480
QY 536 SNLAKGSSANDAVTLEOKAAKPTLNAGATSVTPTEISVDKSGNVTAPTNYGVKTE 595
DB 481 snlakgssandavtleokaaaptlnagatstvptteisvdksngvntaptnygvkte 540
QY 596 LNSDGTSDKFSVKGSGNNLSLVTAEHLASLYNEVNRNTADSALQSFTYKEEDDDANAITY 655
DB 541 lnsdgtstkfsvkgsgnnslvtahlaslynevnrntadsalqsfytykeedddanaity 600
QY 656 AKDTTKNAGAVSIILKKGKNGLTVAATKDDTGTFTGLSDQSGSLTIGKSTLNDGLTVKDTN 715
DB 601 akdtknagavsiilkkgkngltvatkddgtftglsgdsgltigkstlndgltvkdtn 660

QY 716 EQIQVANGIKFTNVNGSNPGTGIANTRIRDRKIGFAGSDGAVDTNKPYPYLDQDKLQVGN 775
DB 661 eqiqvangikftnvngsnpgtgiantaritrdrkigfagsdgavdtnkpyldqdklqvgn 720
QY 776 VKITNTGYNAGKALUTGLSPTLPSTADOSSRNIELGNTIQDKDKSNAASINDILNTGNL 835
DB 721 vkitntgynagkalutglsptlpsadossrnielgntiqdkdkснаasindilntgfnl 780
QY 836 KNNNPIDPFVSPTDIVDFANGNATTATVTHDTANKTSKVYDVNVDDTTHIHJTGDDNKK 895
DB 781 knnnpidpfvsptdivdfangnattatvthdcanstkvydvnnvddtthihjtgddnkk 840
QY 896 LGVKTTLNKTSGANTATNFVNSDEDALVNAKDIAENLTLAKEIHTTKGTADTALQ 955
DB 841 lgvkttlknktsgantatnfvnssdedalvnakdiaenltnlakehttkgtadtalq 900
QY 956 TPTVKKVDENNADANATTVGOKNANNQVNTLTILKGENGLNIKTDKNGTVTFGINTTSG 1015
DB 901 tptvkkvdennadanalvtvgknanqnvtltilkgenlniktngkngtvtfvginttsg 960
QY 1016 LKAGKSTLMDGSLSIKNPPTSQEQIOVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFT 1075
DB 961 lkagkstlmdgslsiknptgseqigvgadgvkfakvnnngvvgagidgttritrdeigft 1020
QY 1076 GTNGSLDKSKPHLSKDGINAGCKKITNTOSGELAQNSHDAVTGGKIYDLKTELENKISST 1135
DB 1021 gtngsldkskphlskdginagckkitntosgelagnshdavggkiydlktelelenkisst 1080
QY 1136 AKTAONSLHEFVSVAOQNNFTVSNPYSYDTSKTSDDITFACENGITTKVANKGVVRVGI 1195
DB 1081 aktaonslhefvsvaeqnnftvsnpyssydtstsdvitfagengitckvankgvvrvgi 1140
QY 1196 DQTKGLTTPKLVGNNGKGIVIDSQNGQNTTIGLSNTLIANTVNTNKGSVRTTEQGNIIKD 1255
DB 1141 dqtkglttpklvgnngkgividsqngqnttigsntlianvntndkgsrvrtteqgnikd 1200
QY 1256 EOKTRAASIVDVLISAGFNLQNGEAVDFVSTVDTVNFADGNATTAKVTYDDTSKTSKVYV 1315
DB 1201 eoktraasivdvlisagfnlqnggeavdftvtnfadgnattakvtyddtsktskvvy 1260
QY 1316 DVNVDTTIEVKDKLIGVKVTTTTSTGTGANKFALSQATGDALVKASDITVAHLNLTSGD 1375
DB 1261 dvnvdttievkdklignvkvttstgtgankfalsnqatgdalvkasdlvahlnltsgd 1320
QY 1376 IQTAKASQANNSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTEKAVDKLVAQAQTP 1435
DB 1321 iqtakasqannsagvydadgnkvlydystdnkyqakndgtvdktekevdklvaqaqcp 1380
QY 1436 DGTLAOMNVKSVINKQVNDANKKQGINEDNAFVKLEKAASDNKTNAAVTVGDLNAYA 1495
DB 1381 dgtlaomnvksvinkqvndankkqginednafvglekaasdnkcknaavtvgdlnaya 1440
QY 1496 QTPLTFRAGDTGTTAKLGETLTIKGGQTDNTNKLTDNNIGVWAGTDFGVKLANDLNLNS 1555
DB 1441 qtpltfagdtgttaklgetltikggqtdntnkltdnnigvwaegtdfgvklakdlnlns 1500
QY 1556 VNAGGTFIKDDKGVSFVDSGQAKANTPVLISANGLDILGGKVISNVGKGTDKTDAANVQQLN 1615
DB 1501 vnaggtkiddkgyvsfvdsqgqakantpvlisangldilggkvisnvkgtktdtadanvqqln 1560
QY 1616 EVRNLLGLGNAGNDADGNQVNIADIKDPNSGSSNRTVIKAGTVLGGKNNNDTEKLAT 1675
DB 1561 evrnllglgnagnadgnqnviadi kdpnsgssnrtvikagtvlggknnndteklat 1620
QY 1676 GGTOVGVDKGNANGDLNSVWVTQDKGSKALLATYNAAGQTNLTNNPAAIDRINEQ 1735
DB 1621 ggtovgvdkgndngdlnsvwvtqdkgskalla tynaagqtnltnnpaeaidrineq 1680
QY 1736 GIRFFHVNDGNQBPVQGRNGIDSSASGKHSVAIGFOAKADGAAVAIGRQTOAGNQSIA 1795
DB 1681 girffhvndgnqbpvqgrngidssasgkhsvalgfoakadgaavaigrtqtoagngsia 1740
QY 1796 IGSNAQATGDSQSIAGTGTGNVNVAGKHSIGAIDPSTVTKADNSYSGVNNNQFTDQTQDFGV 1855

Db 1741 igbnaqatgddsgiaigtgnvvagkhsgaigdpstvkadnsysvgnmqftdatqtdvfgv 1800
Qy 1856 GNNITVTESNSVALGNSAISAGTHAGTQAKKSDCTAGTATTTAGATGTVKGFAGOTAVGA 1915
Db 1801 gnnitvtesnsvalgnsaisagthagtqakksdgtagtttagatgtvkgfagotavga 1860
Qy 1916 VSVGASGERRIONVAGEVSATSDVANGSOLYKATQSTIANATNELDHRTHQENKANA 1975
Db 1861 vsvgasgerrionvagevsatstdvangsolykatqstianatneldhrthqenkana 1920
Qy 1976 GISSAMAMASMPQAYIPGRSMVTGSIATHNGOGAVAVCLSKLSLSDNGQWVFKINGSADTQ 2035
Db 1921 gissamamampqayipgrsmvtgsiathngggavavglslsdngqvwfkingsadtq 1980
Qy 2036 HVGAAGVAGFH 2047
Db 1981 hvgaavgagfhf 1992
RESULT 5
AAB69135
ID AAB69135 standard; Protein; 2053 AA.
XX
AC AAB69135;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.
XX
KW Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
PN WO200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
XX WPI; 2001-159722/16.
DR N-PSDB; AAF59104.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis -
XX
PS Claim 1; Fig 4A-V; 247pp; English.
XX
XX The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain Q8 200kDa protein, which is given in the
CC exemplification of the present invention.
XX
XX Sequence 2053 AA;
SO

Query Match 74.5%; Score 7675; DB 22; Length 2053;
Best Local Similarity 74.8%; Pred. No. 0;
Matches 1604; Conservative 134; Mismatches 219; Indels 186; Gaps 28;
Qy 1 MNHYKVTFNKATGTFMAVEYAKSHSTGGSCATGQVGSVCTLSFARIAALAVLIGAT 60
Db 1 mnhykvtnkatgtfmaaveyakshstggscatgqvgsvrtlsfariaalavligat 60
Qy 61 LSGSAYACKKTHHTAIGEONQPRS--GTAKADGDRAITAIGENANAGGQQAIAIGSSNK 118
Db 61 lngsayagqitk-lelgtgtnkinntlkgdalatgeaslaifgslskagsgaiaigsvkp 119
Qy 119 TVNSSSLKIGTDATGQESIAIGDVKASGASIAIGSDDLHLDDQHGNPKHPKG--TLIN 177
Db 120 dpnngsnngvshakgnesiaigdvlaegdasiaigsddlyl-----pknlldkn 170
Qy 178 D---LINGHAVLKEIRSSKDNNDVKYRRTTASGHASTAVGAMSYAQGHSNAPGTRATAKS 234
Db 171 efhklhghheilkkigtstgdkikyrrtraggghastavgamsyagghfsnafgtataea 230
Qy 235 AYSLAVGLAATAEGOSTIAIGSDATSSSLGALALCAGTGAQLOGSIALQGQSVVTQSDNN 294
Db 231 ayslavgladqatkqsslavgsnakanaafaataiggnkvnlgrgvalfgfsgqilrdnn 290
Qy 295 S-RPAYTPNTQALDPKFOATNNTKAGPL--SIG-----SNSIKRKIINVGAGVKNKTDVNV 347
Db 291 tdasayvplgktladqykatrggdstdfisgnsmnnssirrkliinvagarddavnv 350
Qy 348 AQLEAVVAKWAKERRITFO--GDDNSTDVKIGLDNLTITKGAETNALTDNNNGVKEADN 405
Db 351 aqlkliveela-nrkitfgdgdnnsverglntltikgdaqtnalteanlgvv--tdg 407
Qy 406 SGLKVKLAKTLNNLTENVNTTILNATTVKVGSSSSSTIAELLSDSLTFQTPWNGSSTSKT 465
Db 408 nglkvklakeitgltsv-----satnkitvsnntnnnaelsgggtfts-pitgk-tdkt 460
Qy 466 VYGVNGVRFNTN-----
Db 461 vysidglkftndsnsiatkgttrtkkkigfagtdngvdeskpyldneklkvgnatlng 520
Qy 478 -----AETTAAGTTRITRDKIGFARGDGDVDEKQAP 508
Db 521 sltvnnttgnkqivgangikfatvannvantsatvgtariceekigfagtdngvdeqap 580
Qy 509 YLDKKQLKVGVAITIDNGIDAGNKKISNLAKGSSANDAVTIEQLKAAKPTLNAGANGISV 568
Db 581 yldkerlkvrveitdsginaghnkigtltngiantdavitkikldakptinagdgisi 640
Qy 569 TPTEISVDAKSGNVPTAPTYNIGVKTTTELNSDGT--DKFSVKGSGTNNLSLVAEHLASYL 626
Db 641 nsngdglvdsngnitplyniskvttklnsgtsgnknkfsvenahdnslvtakdladyl 700
Qy 627 NEVNTADSAIQSFVKEEDDDANAITVAKDTTNAGAVSILKLGKNGLTAVATKK-DG 685
Db 701 nkvnetsadspsfkvqngdhsn-naitygkdt--ngktfntlikgengvnttnratg 757
Qy 686 TWTFLGSDQSLGTICKSTLINDGLTVKDTNEIQIOWGANGIKETNVNNGSPGPGTANTARI 745
Db 758 tvtfidgslnglttptkltcvsgd-----tngn-----rl 785
Qy 746 TRDKITGAGSDGAVDTNKPYLDDQLQGVNKTITNTGINAGGKAITGLSPTLPSTADQSS 805
Db 786 vieqv-psadg-----nstknlikglsptlpsiaspsg 817
Qy 806 RNIEGNTIQDKDKSNAASINDILNTGNLKNNNRPIDFVSTYDIDVFANGNATTATVTH 865
Db 818 rnialgntieekdknsaasiddvlnaglnknkgdkdfvstydytdvfdignattatvty 877
Qy 866 DTANKTSKVVDVNVDDTTIHLTGTDNNKLGKVKTKLNKTSANGTATNFNVNSDEDA 925
Db 878 deandtksvaydvnvdektietlgtgdnkgkqlgkvtkiktetstgn-attf--stddda 934

Qy	926	LVNADIAENLNTLAKEIHTTGTADTALQFTTVKVVDDNNADDANAITVCKNANQV	985
Db	935	lvksadiagnlntlaeeihktgtentalqftvkvdendkaddnaitvgkdgtskv	994
Qy	986	NTLTLKAGNCLNTKDKNGTVFVFGINTSGLKAGKS-TLNDGGLSITKPNPSEQIQVGAD	1044
Db	995	ntlikgknngldktdkdgctvtfingntqgglkagdsttlnnglsiktasneqilqvgad	1054
Qy	1045	GVKFAKVNNGVVGAGIDGTTTRTRIDEIGTGTGSLDKSPHLSKDGINAGGKKITNQ	1104
Db	1055	gvkfamv-nngvvvgagldgttrtrideigftgtnslskspkhlksdginaggkkitnlq	1113
Qy	1105	SGEIAQNSHDAVTCGKIYDLKTELENKISTAKTAQNSLHEFSVADEQGNFTVSNPYSS	1164
Db	1114	sgeiaknshdavtggklyldkteleknistaktagnslhefsvadeqgnftvsnpyss	1173
Qy	1165	YDTSKTSDVITFAGENGITTKVKNKVVRVIGIDQTKLTPKLTGVNNGKGIIVDSQNG	1224
Db	1174	ydtksdsvitfagengittkvknkvvrvgidqtkltpkltcvgnngkgivinsqngq	1233
Qy	1225	NTTGLSNTLANVTNKGSVRTTEOGNIKDEDKTRAASIVDVLISAGFNLOGNGEAVDFV	1284
Db	1234	nttfglsntlanvtnkgsrvrtteqgnlikdedktraasivdvlsagfnlogngeavdfr	1293
Qy	1285	STYDTVNFADGNATTAKVTYDDTSKTSKVYDVNVDDTTIEVKDKLGVKTTTLTSTGTG	1344
Db	1294	stydtvnfngntttakvtyddtsktskvvydvnvddttievdkdklgttltstgtlg	1353
Qy	1345	ANKFALSQATGDALVKASDIVAHLNPLSGDIOTAKGASQANNSAGYVDADGNKVYDST	1404
Db	1354	ankfalsnqatgdalvkasdivahlntlsgdigtakgasqannsgyvdadgnkvlydst	1413
Qy	1405	DNKYQAKNDGTVDKTEVAKDKLVAOAQTPDGTLAQMNVKSVINKEOVNDANKKOGINE	1464
Db	1414	dnkyqakndgtvdktevakdklvaqaqtpdgtclaqmnmvksvinkeovndankkagine	1473
Qy	1465	DNAFVGLERKAASDNKTKNAAVTVGDLNVAQAOTPLTFAGDTGTTAKKLGETLTIKGGQTD	1524
Db	1474	dnafvglekaasdnktnaavtvgdlnvavqaotpltfagdtgttakkllgetltlkgqtd	1533
Qy	1525	TNKLTDNNIGVAGTDFGTYKLAKDLNLNSVNAAGTPIIDDKGVSFVDSGQAKANTPVL	1584
Db	1534	tnkltddnigvagtdfgtyklakdlnlnsvnaagtpiiddkgvsfvdsgqakantpvl	1593
Qy	1585	SANGLDLGGKVISNVGKGTDKTDAANYQQLNEVRNLLGLGNAGNADGNQVNIADTKKD	1644
Db	1594	sangldlggkvisnvgkgtktdaanvqqlnevrnllgl---gndnagngvniadlkkd	1650
Qy	1645	PNSGSSNRVTIKAGTVLGGKGNNDTEKLTATGGTQGVGDKDGNANGDLSNVWKTOKDGS	1704
Db	1651	pnsghssnrvtikagtvlggkgnndteklataggvqvgvdkdgnangdlsnvwktqkds	1710
Qy	1705	KKALLATYNAAGQTNLYLNNPAAEATDRINEQIRFFHVNDCNQBPPVQGRNGIDSSASGK	1764
Db	1711	kkallatynaagqtnvlylntnpaaeatdrineqirffhvnndcnqbppvqgrngidssasgk	1770
Qy	1765	HSVAIFQAKADGEAAVAIGRQTAGNQSIAIGDNAQTGQOSTAIGTGNVVAGKHSQAI	1824
Db	1771	hsvalgfqakadgeaavaigrtqtagndqsiaigdnaqtgqdsiaigtgnvvagkhsqai	1830
Qy	1825	GPSTVTKADNSVSVGNNOFTDATQTVFGVGNNTVTESNSVALGSNSAISACTHAGTQ	1884
Db	1831	gpstvtkadnsysvgnnoftdatqtdvfgvgnntlvtesnsvalgsnsaisacthagtcq	1890
Qy	1885	AKKSQCTAGTTTGTAGATGTVKGFAGQATGAVGAVSVGASGAERRIQNVAAGEVSATSTDAVN	1944
Db	1891	akksqctagttttagatgtvkgfagqatgavgvsgasgaerriqnvaagevsatstdavn	1950
Qy	1945	GSOLYKATQSIANATNLDHRTHQENKANAGISSAMAMASMPQAYIPGRSMVTGGTATH	2004
Db	1951	gsolykatqatsianatneldhrihqnkenkanagissamamampqayipgrsmvtggath	2010

2005 NQOGAVAVGLSKLSDNGQWVPKINGSADTQGHVGAAYGAGFHF 2047
|||||

2011 ngqavavglsklsgdngqwwfkingsadtqghvgaavagfhhf 2053
|||||

RESULT 6
AAB69136
ID AAB69136 standard; Protein; 2314 AA.
XX AAB69136;
AC AAB69136;
XX 24-APR-2001 (first entry)
XX M. catarrhalis les1 200kDa protein SEQ ID NO:11.
XX Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX Moraxella catarrhalis.
OS
XX WO200107619-A1.
PN
XX 01-FEB-2001.
PD
XX 26-JUL-2000; 2000WO-CA00870.
PF
XX 27-JUL-1999; 99US-0361619.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
PI
XX WPI; 2001-159722/16.
DR
XX N-PSDB; AAF59105.
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis
PT
XX Claim 1: Fig 5A-Y; 247pp: English.
XX The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis les1 200kDa protein, which is given in the exemplification
CC of the present invention.
XX
SQ Sequence 2314 AA:

Query Match 64.0%; Score 6593.5; DB 22; Length 2314;
Best Local Similarity 60.7%; Pred. NO. 2.3e-282;
Matches 1466; Conservative 149; Mismatches 330; Indels 471; Gaps 48;

Qy 1 MNHYKYVFNKATGTMAVAEYAKSHSTGGSCATGQVGSVCTLSFARTAAALVLVIGAT 60
|||||
Db 1 mnhykyvfnkagtgmavaeacskshsgssstaggvgspsvirtrvatlailvgat 60
|||||
Qy 61 LSGSAYAQKDKTHIAITGEONQPRRSCTAKADGDRATIGENANAGGQAIATGSNKTV 120
|||||
Db 61 lngsayaq--nnsklafgttg---nndnasasneaisalgsalakhangqalggskpdp 115
|||||
Qy 121 -NGSSLDKIGTDTGQESIAIGGDVKAAGSDASIAIGSDDLHLDDQHGNPKHKGTLINDL 179
|||||

Thu Sep 13 14:18:50 2001

```
QY 1972 KANAGISSAMAMASMPQAYIPERSMVTGGIATHNGQGVAVGLSKLSDNGQVWFKINGSA 2031
|||||
Db 2239 kanagissamamaspqayipgrsmvtggiaithngqgvavglsklsdngqvwfkinga 2298
|||||
QY 2032 DTQGHVGAAGAGFHF 2047
|||||
Db 2299 dtqghvgaagagfhf 2314
|||||

RESULT 7
AAR99393
ID AAR99393 standard; Protein; 2353 AA.
XX
AC AAR99393;
XX
DT 15-JAN-1997 (first entry)
XX
DE Haemophilus adhesion protein HA2.
XX
KW Haemophilus adhesion protein; HA2; hsf protein; vaccine.
XX
OS Haemophilus influenzae type b strain C54.
XX
PN WO9630519-A1.
XX
PD 03-OCT-1996.
XX
PF 22-MAR-1996; 96WO-US04031.
XX
PR 24-MAR-1995; 95US-0409955.
XX
PA (UYSL-) UNIV ST LOUIS.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Barenkamp SJ, St Geme JW;
XX
XX WPI; 1996-455364/45.
XX
DR N-PSDB; AAT41476.
XX
XX
XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
PT vaccines against H. influenzae infection.
XX
XX Claim 5; Page 66-73; 120pp; English.
XX
XX Haemophilus adhesion protein HA2 (AAR99393) is associated with the
CC formation of surface fibrils involved in adhesion to various host
CC cells; it is also referred to hsf (Haemophilus surface fibrils).
CC Its amino acid sequence was deduced from a genomic DNA clone
CC (AAT41476) derived from Haemophilus influenzae type b strain C65.
CC Large quantities of recombinant HA2 can be produced in transformed
CC prokaryotic or eukaryotic host cells, for use in vaccines against
CC H. influenzae infection.
XX
XX SQ Sequence 2353 AA;

Query Match 10.9%; Score 1123; DB 17; Length 2353;
Best Local Similarity 23.0%; Pred. No. 1.6e-41;
Matches 607; Conservative 318; Mismatches 838; Indels 874; Gaps 119;

QY 1 MNHIIYFNKATGTFMAVEAYAKSHSTGGSCATGQVGSVCTLSFARIALAV----- 54
|||
Db 1 mnkifnviwmtqtwwvseilrthtk--rasatvetavlatlilfatcvqanadedeel 58
|||||
QY 55 -LVIGATLGSAYAOBKDTKHIIAGEQNPQRS-----GTAKADGDRAIAIGENAN 104
|||
Db 59 dpvvtapvlsfhsdegt-----gekevtenwnwgiyfdnkvla-gaitlkagdnk 112
|||||
QY 105 AQGQGAIAIGSNKVTNVS-----SLDKTGTDTAT--QGESIAIGGDVAKSGD-ASIAIGSD 157
|||||
Db 113 ik-----qntdestnassfysikkldtitsvateklsgf-----angdkvdltsdan 161
|||||
```


Db 1087 kklvnaeglaia-innlswtakadydgesegetdevkagdkvtfkagknlkvkqsek 1145
QY 1054 -----NGVVGAGIDGT-----TRITREICGFTGTNGS-----LDKSKPH-----LS 1089
Db 1146 dftyslqdtlgtltsitlgttngandrntgcvinkdgitlilangaaagtdaasngtislvt 1205
QY 1090 KDGINAGCKKITNIQSG-----ETAQNSHDVATGCKIYDLKTELENKI-----SSTAK 1137
Db 1206 kdgisagkeitnvksalktykdtqntade-tqdekhvaakneavefvgkngatvsak 1264
QY 1138 TAONSLH-----EFSVADQGNFTVSNPYSYVDTSK----- 1169
Db 1265 tdngkhtvidvaeakvgdgldktdgkiklvndldgmlltvdtkgasvaxgefna 1324
QY 1170 -TSDVITFAGENG-----ITTKVKNKGVRVIGIDQTKGLT----- 1202
Db 1325 vttdattagtnanergkvvvksgngatatetdkkvtatgdvavaka.indaatfvkvvendd 1384
QY 1203 -----TPKLT 1207
Db 1385 satidsptddgandalagdkdtlilkagknlkvrkdgnitfalandskvsatvsakls 1444
QY 1208 VGNNNGK-----GIVIDSONGONT-----INGLSNTLANVTNDKGSVRTTEQGNLIK 1254
Db 1445 lgtngknvntsdtkglnfakdsktgddanlhiinglasiitdtllnsgat-tnlgnggit 1503
QY 1255 DEKTRAASIVDLSAGFNLOG-----NGEAVDFVSTYDVTNFDAGNATTAKVTYD- 1305
Db 1504 dnekkraasvkvlnagvnrvgkpasannqvenidfvatydytdvfvgdkdtsvtves 1563
QY 1306 -DPSKTSKVYDVNDVDTTIEVKDKKLGKVTYTTLSG-----TCAN-KFALSNOAT----- 1355
Db 1564 kdngkrtev-----kigatsvikhngklftgkelkdannngvttvtet 1607
QY 1356 -----GDALVKASDIVAHLNLTSGDIOQTAKGSAQNSAGYVD-----ADGNKVIYDS 1403
Db 1608 dgkdegngltakavidavknagvrvtkt-gangqnddfatvasgtnvtfdgnggttaev 1666
QY 1404 TDNKYQAKNDG---TVDKTKEVAK-----DKLVAQA-----QTPDGTIAQMNKSVINKEQ 1452
Db 1667 t-----kandgsitvkynkvkaaglkldgdklvadtvtltvaadgkvtaqn----- 1711
QY 1453 VNDANKKQGINEDNAFVKGLEKAASDNKTKNAAVTGDNLNAVAQPTLTFAGDGTGTTAKKL 1512
Db 1712 -ngdgkk-----fvdasgladalnklswta-tagkegtgevdpsansagg-----evka 1757
QY 1513 GETLTIKGQDTNKLTDNNIGVVAGTGDFTVKLAKDLTNLSV-----NAG-----GTFKI 1563
Db 1758 gdkvtftag-----dnlkqsgkdfyslkkelkdltsvefkddanggtgsestki 1808
QY 1564 DDGVSFVDSGQA-----KANTPVLSSANGLDLGGKVISNVGKGT-----DTDAA 1609
Db 1809 tkdglittpangagaagantantitsvckkgisagknvtnvsvglkkgfdgghtlangtva 1868
QY 1610 NVQOL--NEVRNLLGLGNAGNDN-----ADGNQVNIAD-----IKKDPNSGS----- 1649
Db 1869 dfekhydnaykldtnldekadnnpvtadntaatvdlrglwvisadkttgepnqeyna 1928
QY 1650 ---SSNRTVTKAGT-----VLGG-----KGNNDTEKLANGG-----IQVG 1681
Db 1929 qvrnanevkfsgnginvsgktlgtngtrvitfelakgevrknsneftvknadgsetnlkvvg 1988
QY 1682 -----VDKDGANGDLSNVWVKTK-----DGSKKALLATYNAAGQNTYLTNN 1724
Db 1989 dmyyskedidpatskpmgt---ktekykvengkvvsangsktevtltnkqsg---yytgn 2042
QY 1725 PAEALDRINQOIRFFHVNDGNQBPVVQGRNGIDSSASGKHSVAIGFOAK----- 1774
Db 2043 --qvadalaksgfel-----gladaaeakfaesakdkqlskdkaetv 2084
QY 1775 -ADGEAAVATGROTPACNQSIATGDNAQATCQ-----STAICTGVNVVAKHSGAIG 1825

Db 2085 nahdkvrfanglnctk-----saatvestdangdkvttttfvktdvelpltcqiy---ntdang 2138
QY 1826 DPSTVKADNSYSVGNNNQFTDATQTFEGVGNNTITVESVVALGSNSAISAGTHAGTQA 1885
Db 2139 nkivkxadgkw-----yelnadgtasnkvevtlgnvdangkk-----vv 2176
QY 1886 KKSOGTAGTFTTATGATGTVKGFAGOTAVGAVSVG-----ASGAEERRTONVAAG 1933
Db 2177 kvtenagdkwyytnadgaadkktkgevsndkvtstdekhhvridpnngsqngkvvidnvang 2236
QY 1934 EVSATSTDAVNGSOLY---KATQSTANATNELDRIHONENKANAGISSAMAMASMPQAY 1990
Db 2237 eisatstdaingsqlayavakvtnliagvnnlekvknkvgradagtsalaasqipgat 2296
QY 1991 IPGRSMVTGGIATINGGAVAVGLSKLSDNGQWFKINGSADTQGHVGAAYGAGPHF 2047
Db 2297 mpgksmaiagssyggnglaigvrsisdngkviirlsgttngsqgktgvaagvgvqw 2353
RESULT 8
AAB23860
ID AAB23860 standard; Protein; 2411 AA.
XX AC AAB23860;
XX DT 17-JAN-2001 (first entry)
XX Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
XX non-typeable Haemophilus influenzae; antiinflammatory; auditory;
XX antibacterial; meningitis; epiglottitis; septicemia; otitis media;
XX diagnosis; immunogenic; antigen.
XX Haemophilus influenzae.
XX WO200055191-A2.
XX 21-SEP-2000.
XX 16-MAR-2000; 2000WO-CA00289.
XX 16-MAR-1999; 99US-0268347.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
XX WPI; 2000-618897/59.
XX N-PSDB; AAA92499.
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
XX use as antigens and vaccines and for treating Hemophilus influenzae
XX infection
XX Claim 1; Fig 24; 275pp; English.
XX The present sequence represents a Haemophilus influenzae adhesin
XX (Hia) protein from the type c Haemophilus influenzae strain API.
XX Hia genes and proteins have antiinflammatory, auditory and antibacterial
XX activities, and can be used in the production of a vaccine. An
XX immunogenic composition comprising an Hia gene, a polypeptide encoded
XX by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
XX protection against disease caused by Haemophilus strains in a
XX susceptible host, preferably a human. An Hia protein is useful as an
XX antigen, in immunogenic preparations including vaccines, as a carrier
XX for other immunogens, and in the generation of diagnostic reagents. Hia
XX is useful for treating diseases caused by the infection of Haemophilus
XX influenzae such as meningitis, epiglottitis, septicemia and otitis
XX media. Recombinant production of Hia favours high recovery of the
XX protein compared to the low recovery of native protein from Haemophilus
XX influenzae species. A truncated protein has a significantly higher
XX amount of recovery than a full-length protein.

XX	SQ	Sequence	2411 AA:
		Query Match	10.8%; Score 1109; DB 21; Length 2411;
		Best Local Similarity	23.2%; Pred. No. 6.8e-41;
		Matches	Conservative 306; Mismatches 845; Indels 910; Gaps 124;
Dy	848	giattlpsagakshvdlnvdatkksnaasledvlragwlnqgnngnnvyatydtvn	907
QY	853	FA-----	854
Dy	908	fddstgtttvtqtadgkgadvkligaktsvikdhngklftgdkldkdanngatvaedd	967
QY	855	-----NGNAT	859
Dy	968	kdtgtglvtaktvidavnksgwrvtgegataetgatavnagnaeetvtsgstvfknagnat	1027
QY	860	TANVTHDTANKTSKVVDNVV-----DDTTHLHG-----TDNDK	894
Dy	1028	catskdngn--invdydnvgdglkigdckivadtltitlvlggkvspagansvnnkk	1085
QY	-895	KL-----GVKTTLNKTS-----ANGNT-----AIPFNVSDEED	924
Dy	1086	klvnaegl-atalnlnswtakadyagesegedtdqvkgadvkfkgknlkvkgsqd	1144
QY	925	ALVNAKDI-----	940
Dy	1145	ftyqlgdtlgtlsltggtangrndtgvtnkdgltitlangaaagtdaangntisvtk	1204
QY	941	-----KEIHTTKGTADTALQTFTVKVD-----ENNAD	969
Dy	1205	dgisagnkeltnvklsalktykdtqntagatqaantaevakgdilvdltkpatgaangad	1264
QY	970	-----DANAIVTGO-----KNANNQVNTLTILKGNGLNLI-	998
Dy	1265	akapdttaatvgdlrgwlvsakktadetqdkfhaavkane-----vefvngatavs	1320
QY	999	-KTDKNGTVTFEGINTTSGLKAG-----RSTLNDGGLSIK--NPTSEQIQVGA-DGVREAF	1050
Dy	1321	aktndmgkhtvligvaea-kvgdglekdt--dgkiklkvdndtgmllitvatdkgasvak	1377
QY	1051	VNNNGVV--GAGIDGTTTRITRDEIGFTGTNGS---LDKSHPHLSKDGINAGGGKITNIQ	1104
Dy	1378	gefnavttadatdagtnanergkvvvgksngatatetdk-----kxatv-1422	
QY	1105	GSEIAQNASHDAVTGGKIYDLKTELENTISSAKTAQNSLHFESFVADEOQNFTVSNPYSS	1164
Dy	1423	--gdvaka.indaatf.kvv-----enddsatl-----ddsptddgan-----1456	
QY	1165	YDTSKTSVITP-AGENGITTKVNGV---VRVGIDQTGKLTTPKLTGVGNNGK-----1214	
Dy	1457	-dakagdtltlkagknlkvkrdrknitfalandsvksatvsdkslgtngknknited	1515
QY	1215	----GIVDSONGNT---ITGLSNTLANVTNDKGSVRTTEOGNIKIDBDKTRAASIVDV	1267
Dy	1516	tkglnfakdsktgddanlhngjastltdtlinsgat-tlngngitdnekkraasvkv	1574
QY	1268	LSAGFNLOQ-----NGEAVDFVSTYDVFNFADGNATTAKVTYD--DYKTSKWVYDV	1317
Dy	1575	lnagwnvrgvpasanngvenidfvyatdvdfvsgdkdttsatvteskngkrtey----	1630
QY	1318	NVDDTTIEVKDKKLGKVTTLTSTG---TGAN-FALSNOAT-----GDALVKAS	1363
Dy	1631	-----kidaktsvikdhngklftgkelkdamngvrtvtedtdgkdegnlgvtak	1678
QY	1364	DIWAHLNTLSGDIOTAKGASOANNSAGYD-----ADGNKYIIDSTONKYIQAANDG-	1415
Dy	1679	avidavnkagvrkkt-gangqniddfatvasgtnvtfadngngttaevt-----kandgs	1731
QY	1416	-TVDKTKVEAK-----DKLVAQA---QFPDGTFLAOMNVKSVINKEQVDANKKQGINED	1465
Dy	1732	itvkynvkvvadglikldgdkivadttlvtvadgkvtpapn-----hgdkk-----1775	
QY	1466	NAPVVGLEKAASDNKTKNAAVTVGDLNVAQAQTPLTTFAGDTGTTAKLGETFLTITKGQDPT	1525
Dy	1776	--fvdasgladalink.lswta-tagkegtgevdpanasgg-----evkagdkvtfkag----	1824
QY	1526	NKLTDMNNGVVGAGTGDGFTVKLAKDLNLNSV-----NAG-----GTKIDDKGVSFVDSGG	1576


```
Db 509 gn-daggkatqtlnglnfkfstgdgellnikvendtvtftpkgs-vqgedgkatiq 566
QY 1430 AQAQTPDGTIAQNMVKSVINKEQVNDANKOGINEDNAFVKGLEKAASDNKTNAAVTVG 1489
Db 567 ngkttdglveasel-----veslnklgwkvvgdkdgs-----g 600
QY 1490 DLNVAQPTLTFAGDTGTTAKKLGELTTLTKGGOTDNTKLTDDNNIGVVAGTDGFTVKLAKD 1549
Db 601 eldgasnetlvksgdkvtl--kagenlkvk--qdgcn-----ftyaljde 641
QY 1550 LTNLSV-----NAGGTKIDDRKGVSPVDSG-----QAKANTPVLSANGLDGGKV 1595
Db 642 itgkvsvefdtangangastkktkdgltsitangangaaatadadkikvasdgisagnka 701
QY 1596 ISNVGKGTDTDAANVQOLNEVENLL-----GLGNAGNDNADGNQVNIADIKDPN 1646
Db 702 vknvsglkkfgdanfnltssadnltkqydaykgltndkxgdkqtltvad-----n 756
QY 1647 SGSSNRVIVAKAGTVLGGKGNNDTEKLTATGGIQGVVDKDGANGDLSNV-WVTKQDGSK 1705
Db 757 t-----aatv-----gdrlglgwv-isadktt 777
QY 1706 KALLATYNAAGQNTYLTNPAAIDRINEQGIREFHVNQGNQEPVVOGRNGIDSS---AS 1762
Db 778 gelnkayna-----qvrnanevfkfsg-ngihvsgktvn 810
QY 1763 GKHSVAITGFOAKDGEAAVAIGROTQAGNOSIAGDNOAQTGDSQTAIGTGNVAVACKHSG 1822
Db 811 grrait--felakdena-----lafgyskalrntvaigcgnvvnaeksg 854
QY 1823 AIGDPSTV--KADNSVSGVNNNOFTDATQTDVFGVGN-----1858
Db 855 afgdpnyiedkaggsyafgdnrit-skntfvlngvngaknykangdvdtvtvdkdkdgk 913
QY 1859 ---ITV-----TESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAG 1909
Db 914 ettvtvpkalgatvensvyignk---statckgknlskdsdgtagnnttagttgtnvgfag 970
QY 1910 QTAVGAVSVGASGAERITQNVAAGEVSATSDAVNGSOLYKATQSTIANATNELDHRTHQN 1969
Db 971 atahgavsvgasgeerriqnvaageisatstdaingsqlvavakgvtn-----lagvqnkv 1026
QY 1970 ENKANAGISSNANAMAPQAVIPGRSMVTGGIATHNGQGVAVAVGLSKLSNDNGQWVFING 2029
Db 1027 gkradagtasalaasqlpqasmpgksmvsiaagsyqgnglaigvrsisdngkviirslg 1086
QY 2030 SADTQGHVGAAGVAGGFHF 2047
Db 1087 ttnsqgktgvaagvggqvw 1104

RESULT 10
AAB23859
ID AAB23859 standard; Protein; 1104 AA.
XX
AC AAB23859;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain K22.
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
```

```
PF 16-MAR-2000; 2000WO-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-618897/59.
XX
DR N-PSDB; AAA92498.
XX
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating hemophilus influenzae
PT infection
XX
XX Claim 1; Fig 23; 275pp; English.
PS
XX
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
CC protein from the non-typeable Haemophilus influenzae (NTHi) strain K22.
CC Hia genes and proteins have antiinflammatory, auditory and antibacteri-
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
SQ Sequence 1104 AA;
```

```
Query Match 9.9%; Score 1016; DB 21; Length 1104;
Best Local Similarity 27.9%; Pred. No. 3.4e-37; Indels 394; Gaps 55;
Matches 373; Conservative 166; Mismatches 405;

QY 869 NKTSKVYDVNVDD-----TTIHLTGDNDKKLGKVTTKLNKTS-ANGNTA----- 913
Db 2 nkifvlnwvvtwtvvvseltrahckcasatvavavatlalsataeaanntsvtnglna 61
QY 914 ---TNFN-VNSDEDAVLNNAKDAENLTLAKEIHHTKTGDTALQTFYTKKVDENNAD 969
Db 62 ygdtnfnttnnsiadlekvhqdaykgllnlnek-----dtnksflva----- 104
QY 970 DANAATVVG-----QKNAN-----NQVNTLTLLKGENGLNI-KYDRKNGTVTFGI 1010
Db 105 dntaatvgnlrklgwllsskngtrneksyqvkqadevltfgsgaatvsssskdkhtiti 164
QY 1011 NTTTSLKAG-KSTLNDGGLSIKNPTGSEQIQVGADGVKFAKVNNGVGVGAGIDGTTTRITR 1069
Db 165 svtkgsfaevktdattgg-----qvnadrgkvkaeden---gadvd----- 202
QY 1070 DEIGFTGTNGSLDKSPHLSKDGINAGGKKITNTIOSGEIAQNSHDAVTGGKIYDLKTELE 1129
Db 203 -----kkvatkv---dvakaındaatfvkveestdddie 232
QY 1130 NKISSTAKTAQNSLHEFSVADEQGNFTV---SNPYSSYDTSKTSVDVITFAGENGITTKV 1186
Db 233 ngaagknettdqal-----kagdtitlkagknlkakldqngksvtfalakldivtsak 285
QY 1187 NKGVVRVIGIDQTK-----GLTTPKLTGVNNGNGKIVIDSQNGQNTITGLSNTLANVT 1238
Db 286 vsdklsigkdktnkvdiitsdanglklak--tgngng-----qngnvhlnlgiastl---- 332
QY 1239 NDKGSVFTTEQGNLIKDEKTXRAASIVDVLISAGFNLOGNEAVDFVSTYDTVNFADGNAT 1298
Db 333 tdtlgtmttqasngvavqnhnraasvadvinagwnlqngasvdfvnaaydtvdfvngtnt 392
```



```
QY 1299 TAKVYDDTSKTSKVYVD-----VNVDTTIEVK-----DKKLG--- 1332
Db 393 nvnvtcdtahkktvrvdvtgplpvgyvtedgktvvvdknkyeakqdgdsadmkkvenge 452
QY 1333 -VKTTTLSTGTGANKFALSQAOTDALVKASDIVA--HLNLTSLGDIQTAKGASQANNSA 1389
Db 453 laktvklvlsasgqnpvksisvaeq---teendavsfkqlkalq-ekqvtiltasnayang 508
QY 1390 GYVDADGNKVI-----YSDTNKYYQAK--ND-----GTVDKTKYEVAKDKLV 1429
Db 509 gn-dadgkatgtlnnglnfkfstgdgellnikvendtvtftpkkgsv-qvgedgkatig 566
QY 1430 AQAOTPDCTLAQMNKVSINKEQVNDANKKOGINEDNFAVFKLEKASDNKTKNAATVVG 1489
Db 567 ngtktdglveasel-----veslnklgwkvgvdkdgs-----g 600
QY 1490 DINAVAPLTFAGDTGTTAKKLGELTLIKGGQTDNTLTDNNIGVAVAGTDCFTVKLAKD 1549
Db 601 eldgaasnetlvksqdkvll--kagenlkvk--qdgtl-----ftyalkde 641
QY 1550 LTNLNSV-----NAGGTKIDDKGVSPVDSG-----QAKANTPVLSANGLDLGGKV 1595
Db 642 ltgkvsvefkdtagangastkktkdgittsangangaatdadaikvvasdgisagnka 701
QY 1596 ISNVGKGTDKTDAANVQQLNEVRNLL-----GLGNAGNDNADGNQVNIADIKKDPN 1646
Db 702 vknvsvgiklkgfadanfnltsadnltkqyddyakgllnldeqadkqtlvad-----n 756
QY 1647 SGSSSNRTVIRKAGTVLGGKGNNDTEKLTATGGTQGVDRKDNANGDLSNV-WVKTKDKGSK 1705
Db 757 t-----aatv-----gdrlrglv-isaadktt 777
QY 1706 KALLATYNAAGOTNVLTNPAEADIRINEQGIREFVNDGNOEPVVQGRNGIDSS---AS 1762
Db 778 gelnkeyna-----qvrnanevkfksq-nglhwsgkvtv 810
QY 1763 GKHSVAIGFOAKADGEAAVAIGRTQAGNOSTAIGDQAQATGDQSIATGCTGNVYAGKHSQ 1822
Db 811 grrret---felakdena-----lafgyskairdntvaigtgnvvnnaeksg 854
QY 1823 AIGDPSTV--KADNSYSYGNNNQFTDQTQDVFVGNN----- 1858
Db 855 afgdnyiedkaggsyafgndnrit-skntfvlngvvnakykangdvdtetvtykdkagk 913
QY 1859 ---ITV-----TSENVALGSNSAISACTHAGTQAKKSDGTAGTTTAGATGTVKGRAG 1909
Db 914 ettvtvpkalgatvensylvlgnk---statkdkgnlkdsdgtagnatttagttgtvngfag 970
QY 1910 QTAVGAVSVGASGAEERRIQNVAAAGVSAVSATSDAVNGSOLYKATQSIANATNELDHRHQN 1969
Db 971 atahgavsvgsageerriqnvaageisatdsdaingsglyavakgvtn-----lagqvknv 1026
QY 1970 ENKANAGISSAMAMSPQAYIPGRSMVTGGTATHNGQCAVAVGLSKLSLDNGQWYFKING 2029
Db 1027 gkradagtasalaasqlpaasmpkmsviagssyvggnglaigvrsisdngkviirislq 1086
QY 2030 SADTQGHVGAAGAGCFHF 2047
Db 1087 ttnsggktgvaagvggyq 1104
```

RESULT 11

```
AAB23857
ID AAB23857 standard; Protein; 1004 AA.
XX
AC AAB23857;
XX
DT 17-JAN-2001 (first entry)
DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain M407.
DX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
```

```
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
OS Haemophilus influenzae.
PN WO200055191-A2.
PD 21-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-CA00289.
XX 16-MAR-1999; 99US-0268347.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
XX WPI; 2000-618897/59.
XX N-PSDB; AAA92496.
```

Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Hemophilus influenzae infection

Claim 1; Fig 21; 275pp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia) protein from the non-typeable Haemophilus influenzae (NTHi) strain M407. Hia genes and proteins have antiinflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher amount of recovery than a full-length protein.

XX Sequence 1004 AA;

Query Match 9.4%; Score 970; DB 21; Length 1004;
Best Local Similarity 28.1%; Pred No. 3.2e-35;
Matches 341; Conservative 146; Mismatches 341; Indels 386; Gaps 44;

```
QY 979 KNNANQVNTLTLLKGENGLNIKTDKNGTVTFGINTTSLGKAGKSTLNDGGLSIKNPTGSBQ 1038
Db 28 krasatvatavl-----atiltstvqattggttstnglkaygst----- 67
QY 1039 IQVGADGVKFAKVNNGVVGAGIDGTTTRTRDEIGFTGT-NGSLDKSKPHLSKDGINAGG 1097
Db 68 -----nnpnfnaagsatdlarq---fdgaydgllnlnkdanknllvtdd 110
QY 1098 KKITNIQSGEITAQNSHDAVTGGKIYDLKTELENKISSITAKTAQNSLHFSVADEQGNFT 1157
Db 111 k-----aatvgnlrklglwlsksgntrneksqqvkhadevifegkdgyt 154
QY 1158 VSNPYSSYDTSKTSQD---VITFAGENGITTKVKNKGVVVRVGIDQTKGLTTPKLTVGNN--- 1211
Db 155 v-----tsksengkhtvtftlekdlnvk-----natvskdksilgangnk 193
QY 1212 -----NGKGIVIDSQGNQ---TITGLSNTLIANVTNDKGSVTRTEQGNIIKDEDKTRA 1261
Db 194 vdietsdnglklfakpstngqngvnhngiaastltdtit--gttkaatng--vdvqghnra 249
QY 1262 ASIVDVLGAGFNLCQNGEAVDFVSTYDTVNFADGNATTAKTAKTYDDTSKTSKVYDVND 1321
Db 250 asvadvlmagwniagngasvdfvntdytdvfvngintvnt-tdtahnkkt--vrvdv 306
```


[illegible]

Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine; KW

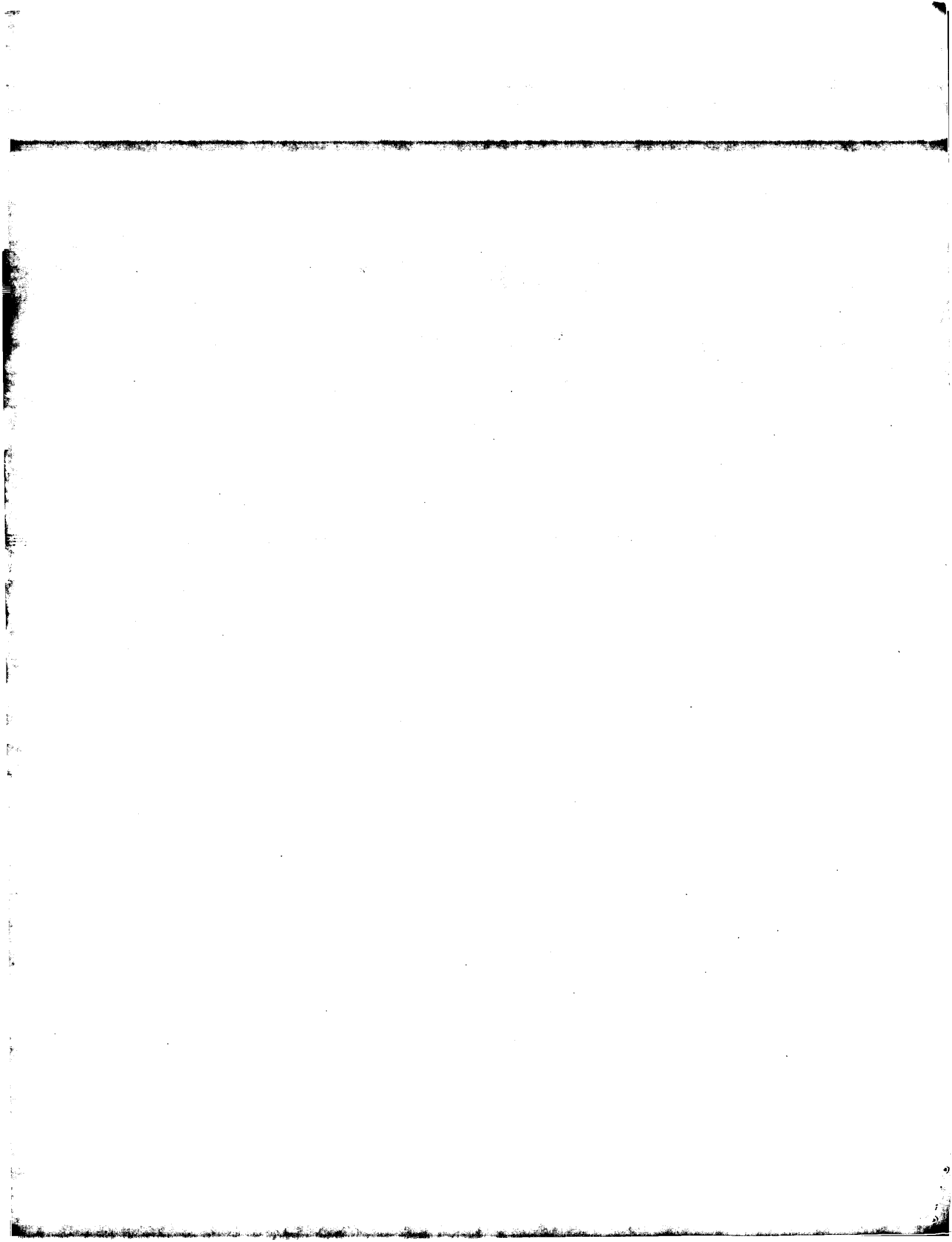

```
Db 316 -----yv 317
Qy 1393 DADGNKVIYDSTDNKYQAKNDGVTDKTEKVAOKLV---AQQTDPDGTLAQMNKSVIN 1449
Db 318 tedsktvv--kvgneyeakqgsadmkkvengklaktkvklvsangt-npvkisinavp 374
Qy 1450 KEQVND-----NKKQGINEDNAFVKLEKASDNKTKNAVTVG-DLNAVAQTP 1498
Db 375 gteddavsfkqlkalqdkqvltisashayang-----gsdadgkatqtlgndlnfkfst 430
Qy 1499 -----LTFAGDTGTAKKLGET-----LTIRGGOTDTNKLTD-----NNIGVVA 1537
Db 431 dsellnikaagdtvftpkpksvqvgdgkatidgakttgllveaseelvdslnklgwkv 490
Qy 1538 G-----TDG-----FTVKLAKDLNLSV----- 1556
Db 491 gvgkdgtgatdgtthdtlvksgkvtlkkagdnlkvkqegtnftvvlrdelegvksvefkd 550
Qy 1557 -----NAGGTKIDDKGVSVFV---DSSGOA---KANTPVLNSGLDGGKVISNVGKGTDT 1606
Db 551 tengangastkitdgtitfpandangaadadkikvasdgisagnkavknvsvg1kff 610
Qy 1607 DAANVQQLNEVRNLLGLCNAGN-----DNADGNQVNIADIKKDPNSGSSNRVTIKACTV 1661
Db 611 gdanfnpl-----ssadnltkydnaykgtlnldekskgkqtpvadnt---aatv 659
Qy 1662 LGGKGNNDTEKLTGGTQGVVDKGNANGDLSNV-WV-KTKDKCSKKALLATYNAAGQT 1718
Db 660 -----gdlrglgwvisadkttgeskeysaqrnanav 691
Qy 1719 NYLTNNPAEADIRNEOGIRFFHYNDGNQEPVVOGRNGIDSSAGKHVSATGFOAKADGE 1778
Db 692 kfkgnginvsgktldngtr-----eitfelakden 722
Qy 1779 AAVAIGROTQAGNOSIATGDNAQATGDSIAIGTGNVAGKHSIGAIGPSTV--KADNSY 1836
Db 723 a-----iafsgskalrntvaigtgnvvnaeksgafgdpnyiedkaggsy 768
Qy 1837 SVGNNOFTDATQTDVFCVGNH-----ITV-----TESN 1865
Db 769 afgndnrit-skntfvlngvgnakykangdvdtetvtvkdkgkettvtpkalgatven 827
Qy 1866 SVALGSNSAISAGTHAGTQAKKSDGTAGTTTATAGTCTVKGFAGQTAVGAVSVCASGNER 1925
Db 828 svyignk---stacdkkgknlksdgtagntttagttvngfagatahgvsvgasgeer 884
Qy 1926 RIQNVAGEVSATSDAVNGSOLKATQSIANATNELDHRHQENKANAGISSAMAMAS 1985
Db 885 riqnvaageisatstdalngsqlyavakgvtn---lagqvnkvgradagtasalaasq 940
Qy 1986 MPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGOWFKINGSADTQGHVGAAGVGF 2045
Db 941 lpgasmsgksmvsilagssyqgsglaigvrsisdngkviirlsgttngsgktgvaagvgy 1000
Qy 2046 HF 2047
Db 1001 qw 1002

RESULT 13
AAW56319
ID AAW56319 standard; Protein; 2042 AA.
XX
AC AAW56319;
XX
XX 19-AUG-1998 (first entry)
XX
DE Haemophilus paragallinarum antigenic protein #1.
XX
KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW vaccine; chicken infectious coryza; CIC; fowl.
XX
OS Haemophilus paragallinarum.
```

```
XX FH Key Location/Qualifiers
FT Peptide 1..70
FT /label= signal
FT Protein 71..2042
FT /note= "antigenic protein"
XX MO9812331-A1.
XX 26-MAR-1998.
XX 12-SEP-1997; 97WO-JP03222.
XX 19-SEP-1996; 96JP-0271408.
XX (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
XX (KAGA ) CEMO-SERO-THERAPEUTIC RES INST.
XX Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX WPI; 1998-230318/20.
XX N-PSDB; AAV22834.
XX Antigenic polypeptide from Haemophilus paragallinarum induces HI
XX antibody production - and is useful for diagnosis of and preparation
XX of vaccines for chicken infectious coryza
XX Claim 1; Page 51-69; 108pp; Japanese.
XX The present sequence represents an antigenic protein derived from
XX Haemophilus paragallinarum strain A-221. The antigenic protein
XX stimulates the production of HI antibodies in fowl. The protein
XX and DNA coding for it can be used in the preparation of vaccines
XX for the prevention of chicken infectious coryza (CIC). The protein
XX and its antibodies can be used in the diagnosis and treatment of CIC.
XX Sequence 2042 AA;
XX
Query Match 7.9%; Score 813; DB 19; Length 2042;
Best Local Similarity 22.6%; Pred. No. 6,1e-28;
Matches 516; Conservative 294; Mismatches 835; Indels 642; Gaps 105;
Qy 17 MAVAEYAKSHSTGGGSCATQGVSVCTLSFARIALAVLVIGATLGSAYAQKDKTKHTA 76
Db 142 iaigygaknestapstvig-----qainrfeksiv 173
Qy 77 IG-----EQNPRRSCTAKADGDRAITAIGENANAQGOAIAIGSSNKTYNGSSLDKIGTDA 132
Db 174 mglhaytql DPR--gtsketrqgsvvigenaksagnqsvsgnswsktnsisigagfta 231
Qy 133 TGOESIAIGGDVAKSGDASIAIGSDDLHLLDQHGPKPKGTL-----INDLINGHA 184
Db 232 egksiaigt-d-kisgtky---ndklpatawngtgvtpknsiwdifselymgkqtngdr 286
Qy 185 VLKEIRRSKDNVDKYR-----RTTASGHASTAVGAM 215
Db 287 ydttrtdpnkpeafykfsdfkgyvntptasptyagkgaialgertlaagemstavgsl 346
Qy 216 SYAQGHFSNAPGTRA-TAKSAY-SLAVGLAATAEQSIIAIGSDATSSSLGAIAGAGTR 273
Db 347 afaladrstamglrsfvakdvvggtaigeesrtfakdsvaigknkteasnagmaygykak 406
Qy 274 AOLQGSIALGQG-----SVVTQS-D 292
Db 407 avgagaiaigtevaagakfnshqtnllqdnayatlknadksdtktnaivtfqgsfd 466
Qy 293 N---NSRPAYTPN-----TQALDPKFQATNNTKAG-----PLSIGNSIKRKIINYG-AGV 339
Db 467 mltnglplvsenetylttsagaikktattdssagggnaiagskttfaskansvalasy 526
Qy 340 NKTDVAVNQALEA-----VVKWAKERRITFGDDNST--DVKIGLDNLT 382
```


Db 527 aladaqnafalqsfsfveassatntitigvsvyakoknsflggtwastlsdrvtvvlgnats 586
 QY 383 IRGGAFTNALTNNICV---VKEANSGLKVKLAKTLNNLFEVNTTTNATTTVKGSSS 439
 Db 587 issgsq-nala---igvvnflgndaassialgmgti-----aksakspdsiaigkea 635
 QY 440 STAEALLSLSLFTQPTNGSQST-----SKTVGVGVKFTNNAETAAIG--TTRIT 490
 Db 636 ridak-dtongtlypqvyvdeetrafrfnessedymrqamalgfnakvrgvgkmetgin 694
 QY 491 RDKIGFARDGDVDEKOAPLDLKK-----QLKV-----GSVAI-----TIDNGIDAG 531
 Db 695 smaigayacatlnstalgvgsktdytweqletdpwsegalsiptsgktgvisvgsgs 754
 QY 532 NKISNLAKGSSANDAVTTEQLKAAP-----TLNAGAGISVTPTE-LSVDKSGNVT 583
 Db 755 errivnlsgssdtdavnvaqiktveerflseilnllqngvgvyisvketnllngsgsrva 814
 QY 584 A-----PTYN--IGVKTTELNSDG-----TSDKFSVKSGSTNNSL-----VTAHILASLYLN 627
 Db 815 sqirkgenyeryvklitqilyldarklneqkfd-----qsnlkitravvqeleeaysg 868
 QY 628 EVNRADSALQSTVKEEDDDANAITVAKDRTKNAGAVSILKLGKNGLTVAATKDGTV 687
 Db 869 elkttasalnqvatiqe-----vttnfdkfnqyktqlen----- 905
 QY 688 TFGLSQDSGLTICKSTLNNDLGT-----VKDNEIOIQVANG-----IKF-----TNVNGSNP 735
 Db 906 --asnadsarnvg-----glcpqalaikannnyindgakqgdsiafgwqaktesgan- 955
 QY 736 GTGIANTARITRDKIGPAGSDGAVDTNKPYLDQDKLVQV-NVKITNTGINAGKAITGLS 794
 Db 956 -nglagkqala---igfqansa-----enaisigtnsdtsmtgavaigkatvta 1002
 QY 795 PTLPSIQDSQSRNIEGNTTQDKDKSNAAINDILLTGFNLKNNNPIDFVSYDIDVEA 854
 Db 1003 gqkpsia-----lg-----qdstvansa-----istspnm--inglifnfa 1038
 QY 855 NGNATTATVTHDTANKTSKVYVDVNVDDTTIHLTGTDDNKKLGKVTTKLNKT--SANGNT 912
 Db 1039 gspetlgvislgtagrerkiv-nvaagdvsgasteaingsqlyatnfmiskvagsksnf 1097
 QY 913 ATNFVNSSDEBALVNAKDAENLNTLAKELHTTKGTADTALQTFVVK-----KVDENN-- 966
 Db 1098 ggnvnlgd--dgtitftnig-----gtqatihdaiannvltkgyllkagdnpt 1144
 QY 967 ----NADDANAITVQGNK--ANNQVNTLTILKGENINIKTDKNGVTTFGINTTSGLK-- 1017
 Db 1145 gnqgqkveignailisatnqwanngnykt-----nlltynsqngtilfgmredpsvkqi 1200
 QY 1018 -AGKSTLNDGSLSTKNPTGSEIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTG 1076
 Db 1201 tag-----tynttgd-----annknqnlntlqgt-----tleatgits 1233
 QY 1077 TNGSLDKSPHLSKXGI--NAGKKITIQSGEITQNSHDAVTGGKIYDLKTEL--ENKI 1132
 Db 1234 svgstnyagfslgadsvtfskggsgtvkissvgsdatdtdaatlkqkvekyrttllvgndi 1293
 QY 1133 SSTAK---TAQNSLHEFSVADEQGNFTFVSNPYSYDTSKTSDDVTTFAGENGITTKVNGK 1189
 Db 1294 taadrsqgtsngitynlsl-----nkgtsvateekvsvgkt---vyealnrait-----gn 1341
 QY 1190 VYRVGIDQTKGLTTPKLVGNNGKGIVIDSQNGQNTITGLSNTLANVTDNKGSVRTEQ 1249
 Db 1342 iftiglddt---tlnk--innpadqdlslseshgnaitg----lvdvvktnspitvpep 1392
 QY 1250 GNIIDEKTRAASIVDVLISAGFNLQNGEAVDFVSTYDVTNFDAGNATTAKVTYDDTSK 1309
 Db 1393 st---dsnkktftv-----gvdfdt-----itegdtdkkltsksv 1429
 QY 1310 TSKVYVDVNVDDTTTEVKDKKLGKVTTLTSTGTGANKFA-----LSNQATGALVKAS 1363
 Db 1430 esyvnkianfstdilldsgrgnatt--andvgkrrlsdgftiksenftlsgkqyngs 1487

QY 1364 DIVAHLNLTSGDI-----QTAKGASQANNSAGYVDADGNKVIYDSTDNKKYIOAKNDGTV 1417
 Db 1488 dslgvyddqngvfklsimlmtalttslant---fakidaenltdsdnskekwrtaln--vy 1542
 QY 1418 DTKT---EVAKDKLVAQAOTPDGTLAQMNKSVYNKEQVNDANKKOGINEDNAFVKGLEK 1474
 Db 1543 sktevdaiqskvvl---tpdegl-----ifatkagsgn-nagldagnkksinv-- 1589
 QY 1475 AASD-NKTKNAAVTVGDLNVAQAOTPLTFAGDTGTAKKLGETLTIKGGOTDTNKLTDNNI 1533
 Db 1590 adgdisptsgdvtrgrqlyalmqkgirvygd-----evaptkqtctaptanp-tatta 1642
 QY 1534 GVAAGTGTFTVKLAKDITNLNSVNAVAGTKIDDKGVSVFDSGGQAKANTPVLSEANGLD--- 1590
 Db 1643 ptasstgwa-----ttant--agg--vapagnvatgdiaptqptlpemtalvddhl 1691
 QY 1591 ---LGGKVISNVGKTGTDAAVQV-----QLN-EVRNLLGLGNAGNDNA-----DGNQV 1636
 Db 1692 avp199sliklbgdhvktlisdahqvgislaqlnisiennlvisgndpekeklaaeqnal 1751
 QY 1637 NIADIKKDPNS-----GSSSNRTVI-----KAGTVLGGK-----GNNDTEKLTATGGIOV 1680
 Db 1752 vitn-kddgnaamvfnnnekmlvisdkeakprvilldgngaltlvgnddq--vtlsakk 1808
 QY 1681 GVDKDNANGDLSNVWVKTKQDGSKALLATYNAAGOTNYLTNPAEAIDRINEQGIREF 1740
 Db 1809 gkdidgn---dlrlsvttter-----nadgglekvetsfatmdd-----gikf- 1849
 QY 1741 HVNDGNQEPVVGVRNGIDSSASGKHSVAIGFAKADGPAVAIGRQTQAGNQSTAIAGDNA 1800
 Db 1850 -----kagdkvl-----nkklnetv 1865
 QY 1801 QATGQOSTAIGTCNVVAGKHSAGTDPSTVADNSYSGVNNNQFTDATQTDVFGVGNNT 1860
 Db 1866 elvgdevnt-----tsitddnkvvslnkkia-----idevkiptndp 1903
 QY 1861 VTESNSVALGNSAITSAGTHAGTQAKSDGTAGTTTAGATGTVKGFAGQAVGAVSVGA 1920
 Db 1904 daqgdsivinnngghagnkvltgvykasd----- 1932
 QY 1921 SGERRIONVAAGEVSATSDAVNGSQLYKATQSTANATNELDRIHONENKANAGISSA 1980
 Db 1933 -----dptsavnrgqlntvldnvgnnfngvqnrgldltresragiaga 1975
 QY 1981 MAMASMPQAIIPGRSMVTGCIATHNGOGAVAVGLSKLSDNGQWVFKINGSADTQGHVGAA 2040
 Db 1976 matasiqnvalp9kttisvgtatfkgenavaigmarsldngkvgirlsgmstsgkdkaa 2035
 QY 2041 VGAGFHF 2047
 Db 2036 msvgfsf 2042
 RESULT 14
 AAW56322
 ID AAW56322 standard; Protein; 2039 AA.
 XX
 AC AAW56322;
 XX
 DT 19-AUG-1998 (first entry)
 XX
 DE Haemophilus paragallinarum antigenic protein #2.
 XX
 KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
 KW vaccine; chicken infectious coryza; CIC; fowl.
 XX
 OS Haemophilus paragallinarum.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..70
 FT /label= signal
 FT Protein 71..2039



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2001, 23:14:44 ; Search time 44.72 Seconds
(without alignments)
942.496 Million.cell updates/sec

Title: US-09-361-619-7
Perfect score: 10303
Sequence: 1 MNHIYKVFNKATGTFMAVA.....NGSADTQGHVGAAGAGPHF 2047

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7380	71.6	2123	4	US-08-968-685A-10
2	1123	10.9	2353	4	US-08-377-155-33
3	1123	10.9	2353	4	US-08-913-942-4
4	792.5	7.7	1912	1	US-08-409-995-4
5	792.5	7.7	1912	3	US-08-685-467-4
6	678.5	6.6	1098	1	US-08-409-995-2
7	678.5	6.6	1098	3	US-08-685-467-2
8	678.5	6.6	1098	4	US-08-377-155-32
9	678.5	6.6	1098	4	US-08-913-942-2
10	573	5.6	679	4	US-08-913-942-15
11	517.5	5.0	1600	2	US-08-617-697-10
12	503.5	4.9	1599	2	US-08-617-697-9
13	501.5	4.9	1529	2	US-08-728-470-10
14	501.5	4.9	1529	4	US-08-719-641-10
15	457	4.4	1338	2	US-08-728-470-9
16	457	4.4	1338	4	US-08-719-641-9
17	435	4.2	1612	1	US-08-169-927-2
18	433	4.2	1536	1	US-08-038-682-2
19	433	4.2	1536	1	US-08-302-832-2
20	433	4.2	1536	2	US-08-530-198-2
21	433	4.2	1536	2	US-08-469-880-2
22	433	4.2	1536	2	US-08-728-470-2
23	433	4.2	1536	2	US-08-617-697-2
24	433	4.2	1536	4	US-08-719-641-2
25	418.5	4.1	1477	1	US-08-038-682-4
26	418.5	4.1	1477	1	US-08-302-832-4
27	418.5	4.1	1477	2	US-08-530-198-4

28	418.5	4.1	1477	2	US-08-469-880-4	Sequence 4, Appli
29	418.5	4.1	1477	2	US-08-728-470-4	Sequence 4, Appli
30	418.5	4.1	1477	2	US-08-617-697-4	Sequence 4, Appli
31	418.5	4.1	1477	4	US-08-719-641-4	Sequence 4, Appli
32	398.5	3.9	599	4	US-09-377-155-15	Sequence 15, Appli
33	393.5	3.8	594	4	US-09-377-155-9	Sequence 9, Appli
34	392.5	3.8	598	4	US-09-377-155-5	Sequence 5, Appli
35	391.5	3.8	598	4	US-09-377-155-13	Sequence 13, Appli
36	384	3.7	589	4	US-09-377-155-19	Sequence 19, Appli
37	383.5	3.7	592	4	US-09-377-155-17	Sequence 17, Appli
38	383.5	3.7	594	4	US-09-377-155-7	Sequence 7, Appli
39	378	3.7	592	4	US-09-377-155-2	Sequence 2, Appli
40	375	3.6	591	4	US-09-377-155-11	Sequence 11, Appli
41	375	3.6	591	4	US-09-377-155-21	Sequence 21, Appli
42	368.5	3.6	658	1	US-08-409-995-5	Sequence 5, Appli
43	368.5	3.6	658	3	US-08-685-467-5	Sequence 5, Appli
44	368.5	3.6	658	4	US-08-913-942-5	Sequence 5, Appli
45	343.5	3.3	607	1	US-08-409-995-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSTLA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968 685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 71.6%; Score 7380; DB 4; Length 2123;
Best Local Similarity 68.3%; Pred. No. 0;
Matches 1555; Conservative 118; Mismatches 219; Indels 386; Gaps 20;
QY 1 MNHIYKVFNKATGTFMAVAEYAKSHSTGGSCATGQVGSCTLSFARIAALVIGAT 60
|||||

Db 1 MNHIVKVFENKATGTFMAVAEYAKSHSTGGSCATQGVGSRVTLTSPARIALAVALVIGAT 60
 Qy 61 LSGSAYAKKOTKHIAIGBQONPRSGCTAKADGDRAIAIGENANAGGGAIAIGSSNKTIV 120
 Db 61 LNC SAYA-----GIGISEADGGKGGANARCDKSIATGIDIAQALGSGOSIAIGDNKIVH 112
 Qy 121 NGSSLDKIGTATGOESIAIGDVKRAGSDASTAIGSDDLHLDDQHGNPKHKGKTLINDL- 179
 Db 113 NSNNANICAKASGNESIAIGDVLASGHASIAIGSDLLY-----KKEVQOISELL 165
 Qy 180 --INGHAVLKEIRSSKNDV-KYRRTTASGHASTAVGAMSYAOGHFSNAGFTRATAKSA 236
 Db 166 PIIRGOKALNDIYQALADTLQYRTHAQGHASTAVGAMSYAKGHFSNAGFTRATAECTY 225
 Qy 237 SLAVGLAATAEGOSTAIGSDATSSLSIAIALCAGCPRAOGLSIALGOGSVVYQSDNNSR 296
 Db 226 SUAVGLTATAKAASSIAGVSGAQAIGFAATAVGGSTOVNLRGIALGFGSQVLQKNDVN 285
 Qy 297 P-----AYTP-NTQALDPKFAATNNTKAGPL-----SIGNSIKRKIIINYGAGVKNKTDAYN 346
 Db 286 AANVRAYAPDNPQIDNRYKATFKNGATDVESIGNSNGNDSIRRKLIINVAGSAGSDTDAVN 345
 Qy 347 VAOLEAVVWAKERRITFGODNSTDVYKGLDNTLTIKGAETNALTDNNIGVWKEADNS 406
 Db 346 VAOLKEAVRLA-NRQITFGDDSNRVERKGLKTLTITGGAQTSALTDHNIQVQNGD-- 402
 Qy 407 GLKVKLAKTLNLTETVNTT'LNATT'VVKVSSSTTAEILLSLSTFTOPTGTSQSTSKTV 466
 Db 403 GLKVLQAEHLTSLUKWTTENLTANEKVYVK----- 433
 Qy 467 YGVNGVKTNNAEATTAAGTTRITRDKIGFARDGVDEKOAPYLDKKOLKVGSAVATION 526
 Db 434 -----TRUTDKIGFTNDMNGIDESKPYLDK-----DT 461
 Qy 527 GIDAGNKKISNLAKGSANDATVLEOLKAAPTLNAGAGISVTPTEISVDKSGNVTAPT 586
 Db 462 GIHAGGOKITKLTAGVDDDAATYGLK----- 489
 Qy 587 YNIGVKT'TELNSDGTSDKFSVKSGTGNLSLVTAEHLASLYNEVNRADSALQSFYKE-- 644
 Db 490 -----KVNTAESALQFTTVKKVD 508
 Qy 645 ---EDDDANAITVAKDTTKNAGAVSILKLGKNGLTATVATKKDGTTFGLSQSDSLTIGK 701
 Db 509 KNGNDANSKIITVGNKKNKPDGTQVNT'KLKGENGVDDVTETNGTTFGLNQNGLTUVGN 568
 Qy 702 STLNDGLTVKOT--NEQIOVGANGIKFTNVNGSNPOTGIANTARITR----- 747
 Db 569 STLNDGLSVKNTNSKQIOVGADGITFTDISNKPFGAGIENTRITRDOGIGFANNTGSL 628
 Qy 748 ----- 747
 Db 629 DANKPRLTPTGINAGGKELTNVQSAINPATNGGQDPMNRLSTANTEKSGSAATIKDLYN 698
 Qy 748 ----- 747
 Db 689 LSQVPLTFAAGDTGNVTKLGEILKVKGGKTTADDLTKNIGVVADSTDNSLTVKLAKTL 748
 Qy 748 ----- 747
 Db 749 SLDLAVNTKLTASDKVTVDGNNNTAKLQNGDLTFSKONTGATPATNSKTIYGVYDGLKFT 808
 Qy 748 -----DKIGFAGSGDGVNTKPYLDQDKLOVGNKJINTGINAGGKAITG 792
 Db 809 DNNGIALDGTYYIT'KDKVGFAGQDGLSKPYLDKDLKVKGEVIELTNGINAGGKAITG 868
 Qy 793 LSP'LTSPADQSSRNIELGNTIQDKRKSNAASINDIILNTGFNLKNNNPIDFVSTYDIDV 852
 Db 869 LSNLT'DATNATTCVTLQGLIGVSDTDTTRAASIGDVLNAGFNKNGDKDAFVSTYDIDV 928
 Qy 853 FANGNATATVTHDTANKTSKVYDVVNDVDTT'IHLTGTDDNK-KLGKVT'TKLNKITSANGN 911
 Db 929 FINGNATTAKVTYD--GKASKAVAYDVAVDGT'IHLTGADGNKNGIKGVKTTTLTKTDKAGD 986

Qy 912 TATNFNVNSDEEDALVNAKDIAENLNTLAKETHTTKGTADTALQTFVVKKYVDENNAD-D 970
 Db 987 KAINFSVNSGDDKALINAKDIADNLNTLAGEIRNTKGTADTALQTFQVKVYKENGDDND 1046
 Qy 971 ANAITVQKNANNQVNTLTKGENGLENLTKDKNGTIVTGTINTTSGLKAG-KSTLNDGGLS 1029
 Db 1047 ADTIVGVKDAKTNOVNTLKLKGNGLDIO'NKDGTIVTGTINTTSGLKAGNNTTLLNNLS 1106
 Qy 1030 IKNPTGSEIOVGADGVKFAKVNNGNVVGVAGIDGTTRITRDEIRGTGTNGSLDKSKPHLS 1089
 Db 1107 IKNTAGNEQIOVGADGVKFAKVNNGVVGVAGIDGTTRITRDEIRGTGTNGSLDKSKPHLS 1165
 Qy 1090 KGINAGGKKTINIQSGBIAQNSHDVATGGKIYDULKTELENKISSTAKTAQNSLHEFSVA 1149
 Db 1166 KGINAGGKKTINIQSGBIAQNSHDVATGGKIYDULKTELENKISSTAKTAQNSLHEFSVA 1225
 Qy 1150 DEQGNFTVSNPYSYDTSKTSVLTFAENGITTKVKNKVVRVGVIGDQTKGLTTPKLVG 1209
 Db 1226 DEQGNFTVSNPYSYDTSKTSVLTFAENGITTKVKNKVVRVGVIGDQTKGLTTPKLVG 1285
 Qy 1210 NNGKGIVIDSONGQNTITGLSNTLANVTNNDKGSVRRTEQGNIIKDEDKTRAASIVDVL 1269
 Db 1286 NNGKGIVIDSONGQNTITGLSNTLANVTNNDKGSVRRTEQGNIIKDEDKTRAASIVDVL 1345
 Qy 1270 AGFNLOQNGEAVDFVSTYDVTNPFADGNATTAKVTDYDTSKTSKVYDVNDVDTTIEVKDK 1329
 Db 1346 AGFNLOQNGEAVDFVSTYDVTNPFADGNATTAKVTDYDTSKTSKVYDVNDVDTTIEVKDK 1405
 Qy 1330 KLGVTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIOTAKGASQANNSA 1389
 Db 1406 KLGVTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIOTAKGASQANNSA 1465
 Qy 1390 GYVDADGNKVIYDSTDNKYIQA'KNDGTVDKTEKAVDKLVQAQTPDGTG'LAQNMVKSIN 1449
 Db 1466 GYVDADGNKVIYDSTDNKYIQA'KNDGTVDKTEKAVDKLVQAQTPDGTG'LAQNMVKSIN 1525
 Qy 1450 KEQVNDANKKOGINEDNAEFVKGLEKASDNKT'KNAATVVDGLNVAQTPLT'FAGDGTGTTA 1509
 Db 1526 KEQVNDANKKOGINEDNAEFVKGLEKASDNKT'KNAATVVDGLNVAQTPLT'FAGDGTGTTA 1585
 Qy 1510 KKLGETLTITKGOTD'TNKLTDNNIGVWAGTGT'VKLAKDLTLNLSVNAAGGTKIDDKGYS 1569
 Db 1586 KKLGETLTITKGOTD'TNKLTDNNIGVWAGTGT'VKLAKDLTLNLSVNAAGGTKIDDKGYS 1645
 Qy 1570 FVDSGQAKANTPVL'SANGLDLGGKVISNVGKGT'KTD'AAANYQOLNEVRNLLGLGNAGND 1629
 Db 1646 FVDSGQAKANTPVL'SANGLDLGGKVISNVGKGT'KTD'AAANYQOLNEVRNLLGLGNAGND 1705
 Qy 1630 NADGNOVNTADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLTATGGIQGVYDKDGNAN 1689
 Db 1706 NADGNOVNTADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLTATGGIQGVYDKDGNAN 1765
 Qy 1690 GDLNSVWVTKQDKGSKKALLATYNAAGQNYLTNNP'AEIDRINEQIRFFHVDNGQEP 1749
 Db 1766 GDLNSVWVTKQDKGSKKALLATYNAAGQNYLTNNP'AEIDRINEQIRFFHVDNGQEP 1825
 Qy 1750 VVOGRNGIDSSASGKHSVAIGFOAKADGGA'VAIGRQTOAGNOSIAIGDNAQATGQOSTA 1809
 Db 1826 VVOGRNGIDSSASGKHSVAIGFOAKADGGA'VAIGRQTOAGNOSIAIGDNAQATGQOSTA 1885
 Qy 1810 IGTGNVAGKHSVAIGDPSTVKADNSYSVGNNNQFTDATQTDVFGVGNNTT'VTSNSVAL 1869
 Db 1886 IGTGNVAGKHSVAIGDPSTVKADNSYSVGNNNQFTDATQTDVFGVGNNTT'VTSNSVAL 1945
 Qy 1870 GSN'ALSAGTHAGTQAKKSDGTAGTTTAGATCTGVKGFAGTAVGAVSVGASCAERRIQ 1929
 Db 1946 GSN'ALSAGTHAGTQAKKSDGTAGTTTAGATCTGVKGFAGTAVGAVSVGASCAERRIQ 2005
 Qy 1930 VAAGEVSATSDAVNGSOLYKATQSIANATN'ELDHRT'HQENKANAGISSAMAMASMPQA 1989
 Db 2006 VAAGEVSATSDAVNGSOLYKATQSIANATN'ELDHRT'HQENKANAGISSAMAMASMPQA 2065


```

QY 1255 DEDKTRAAISVDVLSAGFNLOG-----NGEAVDFVSTVTVNFADGNATTAKVTVD- 1305
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1504 DNEKRAASVKDVLNAGWVRGVKSPASANNQVENIDFVATYDITVDFVSGDKDTSVTVES 1563
QY 1306 -DTSKTSKVVYDVNVDDTTIEVKDKKLGKVTTLTSTG-----TGAN-RFALSNOAT----- 1355
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1564 KONGKKEV-----KIGAKTSVIKDHNGKLFTEGKELDANNNGVTVTET 1607
QY 1356 -----GDALVKASDIVAHLNLTLSGDIQTAKGASQANNSAGYVD-----ADGNKVIYDS 1403
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1608 DGRDEGNGLVTAKAVIDAVNKGAWRKTT-GANGQNDPFPATVASTNTVTFADGNGTTEAV 1666
QY 1404 TDNKYYQAKNDG--TVDKTKEVAK-----DKLVAQA---QTPDGTFLAOMNVKSVINKEQ 1452
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1667 T-----KANDGSITVYKVNKVDGLKLDGDKIVADTVTLTVADGKVTPN----- 1711
QY 1453 VNDANKKQGINEDNAFVKLEKAASDNKTKNAAVTVGDLNNAVAQTPLETFAGDTGTAKKL 1512
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1712 -NGDGKK-----FVDASGLADALNKLKSWTA-TAGKEGTGEVDPANSAGO-----EVKA 1757
QY 1513 GETLTTKGGQTDNKLTDNNIGVWAGTDFGVKLAIDLNLNSV-----NAG-----GTKI 1563
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1758 GDKVTFKAG-----DNLKTKQSGKDFYSLKKELKDLTSVEFKDANGGTGSESTKI 1808
QY 1564 DDKGVSPVDSGQA-----KANTPVLKSANGLDLGGKVI SNVKGTK-----DTDA 1609
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1809 TKGLTITPANGAGAGANTANTISVTKDGISAGKAVINNVYSLGKFGDGHGTLANGTVA 1868
QY 1610 NVQOL--NEVRNLLGLGNAGNDN-----ADGNQVNIAD-----IKKDPNSGS----- 1649
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1869 DFEKHYDNAYKDLTNLDEKADNPTVADNTAATVGLRGLGWIVISADKTTGEPNQEYNA 1928
QY 1650 ---SSNRTVTKAGT-----VLGG-----KGNNDTEKLTAGG-----IQVG 1681
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1929 QVRNANEVKEFGNGINGINVSCKTLNGRVTITFELAKGEVYKSNFTVKNADGSETNLVKVG 1988
QY 1682 -----VDKDGANGDLSNVWVKTKQ-----DGSKKALLATYNAAGOTNLTNN 1724
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1989 DMYSKEDIDPATSKPMTG---KTEKYKVENGKVVVSANGSKTEVTTLTKGSG---YVTGN 2042
QY 1725 PAEAI DRINEGIRFFHVNDGNQEPVVGRRNGIDSSAGKHSVAIGFOAK----- 1774
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2043 --QVADAIAKSGFEL-----GLADAAEAEKAFSAESAKDKQSKDKAETV 2084
QY 1775 -ADGEAAVAIGRTOAQAGNOSIAIGDQAATGDQ-----STAIGTGNVYVAGKHSAGIG 1825
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2085 NAHDKVRFANGLNTKV---SAATVESTDANGDKVTTTFVKTVDVELPLTQIY---NTDANG 2138
QY 1826 DPSTVKADNSYSGVNNQFTDATQTDVFGVCNNITVTESNSVALGSGNSALSAGTHAGTQA 1885
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2139 NKIVKADGKW-----YELNADGTASNKEVTILGNVDANGKK-----VV 2176
QY 1886 KSDGTACTTTTAGATGTVKGFAGQAVGAVSVG-----ASGAERIRIQNYAAG 1933
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2177 KVTENGADKWTYTNADGAADTKGEVSNKYSTDEKHVVRVLDPNQNSGKGVVLDNVANG 2236
QY 1934 EVSATSTDAVNGSOLY-----KATQSIANTATNELHRIHONENKANAGISSAMAMSPQAY 1990
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2237 EISATSTDAINGSOLYAYAKGVNLAGOVNNEGKVNKVGKRADAGTASALASOLPOAT 2296
QY 1991 IPGRSMVTVGLATHNGQCAVAVGLSKLSDNGQWFKLINGSADTQGHVGAAGAFHF 2047
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2297 MPGKSWAIAGSSYGQGLAIGVSRISDNGKVIIRLSGTTNSGKGTGVAAGVGYQW 2353

```

RESULT 3

US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS

```

; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A. 39,054
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: A-61053-1/RPT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-913-942-4

```

Query Match 10.9%; Score 1123; DB 4; Length 2353;

Best Local Similarity 23.0%; Pred. NO. 1.5e-60;

Matches 607; Conservative 318; Mismatches 838; Indels 874; Gaps 119;

```

QY 1 MNHIVKVIENKATGTFMAVEYAKASHSTGGSCATGQVGSVCVTLGFARIAALAV----- 54
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1 MNKIFVNIWVMTQTWVWVSELTRHTK--RASATVETAVLATLLFATVQANATDEDEL 58
QY 55 -LVIGATLSCSAYAAOKKDKHTAIGBQNPQRS-----GTAKADGDRAIAGENAN 104
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 59 DPVVRTAPVLSFHSDEKGT-----GEKEVTENSNGIYFDNKGVLKA-GAITLKAGDNILK 112
QY 105 AQGGQAIATGSSNKTNGS-----SLDKIGTDAT--QGESIAIGGDVKASGD-ASIAIGSD 157
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 113 IK-----ONTDESINASSFTYSLKKDLDTLSVATEKLSFG-----ANGDKVDITSDAN 161
QY 158 DLHLLDQHGPKPKPG--TLINDLINGHVAVLKEIRSKDNVDKYRRT----- 202
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 162 GLKLAKTGNVHLNGLDSTLPDAVINTGVLS--SSFTPDNDVEKTRATVTKVQVNLNAGWNI 220
QY 203 ----TASGHASTAVGAMSYAQGHFSNAPGTRATAKSAYSLAVCLAATAGSOSTIAIGSDA 258
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 221 KGAKTAGGNVESVDLSAYNNVEFIT--GDKN-----LDVVLTAKENGKTTVEVFTPK 272
QY 259 TSSSLGAIAGAGTQAQLOGSIALGOGSVVTOQSDNNRSRPAYTPNTQALDPKFOATNNTKA 318
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 273 TS-----VIREKDGKLFTEGKENDNTNKTNSNT-----ATDNTDE 306
QY 319 GPLSIGSNSIKRKIIINVAGVKNKTDVAVNAQLEAVYKWKAKERRITFOGDDNST----- 371
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 307 GNLVTTAKAVIDAVNKGAWRKVTKTTANGQNGDFATV--ASGTNVTFESGDTTASVTKDT 364
QY 372 -----DVKIGLDNT-LTIKGG----- 386

```


RESULT 4
US-08-409-995-4
Sequence 4, Application US/08409995
Patent No. 5646259
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/REF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

Query Match 7.7%; Score 792.5; DB 1; Length 1912;
Best Local Similarity 22.7%; Pred. No. 2.3e-40;
Matches 474; Conservative 239; Mismatches 652; Indels 727; Gaps 99;

Qy 1 MNHYKVFNKATGTFMAVAYAKSHSTGGSCATGOVGS-----VCTLSFARIAALAV- 54
Db 1 MNKIFNVNVMQTWVVVSELTRTH-----TKRLNRGDPVLATLLFATVQANATD 52
Qy 55 -----LVIGATLSSGSAOKKD/KHIAIGRONQPRRS-----GTAKADGDRAIAI 99
Db 53 EDEELDPVVPAPVLVSFHSDEGT-----GEKEVTENSWGIYFDNKGVLKA-GAITLKA 106
Qy 100 GRNAAOQGOAIAIGSNKNTVGS-----SLDKIGTADT--GOESIAIGGDVKAAGD-ASI 152
Db 107 GDNLKKX-----QXTDEXTNASSFTYSLKKDLTDLTSVATEKLSEF-----ANGDKVDI 155
Qy 153 AIGSDDLHLLOHGNPKIPKG--TLINDLINGHVALEIRSSKDNVRYRT----- 202
Db 156 TSDANGLAKTNGNHNGLDSTPLPAVNTNTGVLS--SFTPDNVEKTRAIVKDVLN 214
Qy 203 -----TASGHASTAVGAMSYAQGHFSNAGFTRAITAKSAYSIAVGLAATAEQSTIA 253
Db 215 AGWNINKAGTAGGNVESDLSAYNNVEFIT--GDKN-----LDVVLTAKENKKITEV 266
Qy 254 IGSDATSSSLGAIAGCAGTRAQGLSIALGQGSVVYTSQSDNNSRPAYTPNTQALPKFOAT 313
Db 267 KFTPKTS-----VIKEKDGKLFTEKENNDTNKVTNT-----AT 300
Qy 314 NNTKAGPLSIGNSIKRKIIINVAGVNTKDVAVNAQLEAVVAKKERRITTFQGDNDNST-- 371

Db 301 DNTDEGNGLVTAKAVIDAVNKAQWRVKTITTTANGQNGDFATV--ASGNTVTFESGDGTAS 358
Qy 372 -----DVKIGLDNT-LTIKGG----- 386
Db 359 VTKDTNGNGIIVKYDAKVGDOGDKFDSKKIIVADTALTALTIVTGGKVAETAKEDEDDKKKLVNAG 418
Qy 387 -----AETNALTDNNI-GVVKAEAD-----NSGLKVK-----LA 413
Db 419 DLVPTALGNLSWKAKEADTDGALEGISKDQEVKAGETVTFKAGKNLKVQDCANFTVSLQ 478
Qy 414 KTLNMLEVTNTTTLNATTTVKVSSSTTABELLSLFTPTOPNTGSSQSTSKTV-YGVNGV 472
Db 479 DALTGLTSLI--TLGGTTN--GGNDAKTV-INKDGLTITPAGNGGTGTGTITSVTKDGI 531
Qy 473 KFTNNAETTAIG-----TTRITRDKIGFARDG--DVDEKQAPVLDKQKL 515
Db 532 KAGNKAITNVASGLRAYDDANFDVLNNSATDLNR-HVEDAYKGLLNLNEKNA--NKQPL 587
Qy 516 KVGSAITIDN-----GIDAGNKKISNLAKGSSANDAVTIEQLKAAPKPTLNAGAGIS 567
Db 588 VTDSTAATVGDRLKLGWVSTKNGTKESN-----QVKQADEVLTGAGAA 633
Qy 568 VTPFISYDAKSGNVTAPTYNIGVKTTE-----LNSDGTSDKFSVKSGTNNSLVTAEHLA 623
Db 634 -----TVTSKSEN-GKHTITVSAETKADCGLEKDGDTIKLVNDNQNTDNLVTGN--- 683
Qy 624 SYLNEVNRITADSALQSPVYKE-EDDDANALITVAKDTTKNAGAVSILKLKNGKGLTVATK 682
Db 684 -----NGTAVTKGGFEIVKTGTATDADRGKTV-KDATANDADKKVATVK-----DVA 731
Qy 683 KQGTVTF-----GLSDSGLTIGKSLNNDGLTVKDT-----NEQIOVCANGIKFT- 728
Db 732 INSAATFVKTENLTTSIDEDNPTDNCK-----DDALKAGDTLTFKAGKNLKVARDCKNITF 787
Qy 729 -----NVNGSNPGTGIANTARI-----TRDKIGFAGSDGAVDTNKPYLD 767
Db 788 DIAKNEVKTAKVSDTLTIGGNTPTGGTTATPKVNITSTADGLNFA----- 833
Qy 768 QDKLOVGNVKITNTGINAGGKAI--TGLSPTL--PSIADQSSRNIELGNTIODKDSNAA 823
Db 834 -----KETADASGKNVYLKGIATTLTEPSAGAKSS-HVDL--NVDATKKSNA 879
Qy 824 SINDILNTGFLNKNNPIDFVSTYDIDVFANGNATATVTHDTANKTSKVVDVNVDDT 883
Db 880 SIEDVLRAGNIQGNNGNVDYVATYDTVNFDDSTGTTTIVT-----V 921
Qy 884 TIHLTGTDDNKKLGKVTTKLNKTSANGNTAT-----NFNVNSDED-----ALVNAK 930
Db 922 TOKADGKADVKIGAKTSVIK--DHNGKLTGDKLDANNGATVSEDDGKDTGTGLVTAK 979
Qy 931 DIAENLNTLAKEIHTTKGTADTALQTFVVKVDENNADANAITVGO---KNANNQVN 986
Db 980 TVIDAVNKSGWRVTGEGATAETGATAV-----NAGNAETVTSVNFKNGN--AT 1028
Qy 987 TLLKKGNG--LNITDKNGTVTFEINTTSLKAG-----KSTLNDGGLSLKNPTGSE 1037
Db 1029 TATVSKDNGNINKYD-----VNVGDLKIGDDKIVADTTTLTVTGGKVSVPAGAN 1080
Qy 1038 QIQ-----VGADGVKFAKVN----- 1053
Db 1081 SVNNKKLVNAEGLATA-LNNLSWTAKADKYADGESEGTQDEVKAGDKVTFKAGKNLV 1139
Qy 1054 -----NCVVGAGIDGT-----TRTRDEIGTGTNGS---LDKSKPH 1087
Db 1140 KQSEKDPYSLQDTLTGLTSITLGGTANGRNDGTIVLNKDLGITLANGAAAGTDSNGN 1199
Qy 1088 ---LSKDGINAGKKTINIOG---EIAONSHADVTTGGKIYDLKTELENKI----- 1132
Db 1200 TISVTKDGISAGNKEITNVKSALKTYKDTQNTADE-TQDKEFHAAVKNAEVEFVGKNGA 1258
Qy 1133 SSTAKTAQNSLIH-----EFSVADEQGNFTVSNPYSSVDTSK 1169

Db	1259	TWSAKTDNNGKHTVTDVAEAKVGDEKEDTDGKILKLVNDTGNLLTVDTATKGASVAK	13118
QY	1170	-----TSDVIITFAGENG-----ITTKYKGVVRVIGDITQKGLT-----	1202
Db	1319	GEFNAVTTDATTAQGTNANERKGVVVKSGNGATATETDKKKVATVGDVAKAINDAATEVVK	1378
QY	1203	-----	1202
Db	1379	VENDSATIDDSPTDDGANDALAKAXDTITLAKAGNLKVKRDKGNITFALANDLSVKSATV	1438
QY	1203	TPKLTIVGNNGK-----GIIVDSQNGONT-----ITGLSNITLANVNDKXGSRVITEQ	1249
Db	1439	SOKLSLGTNGKNVNITSDTKGLNFAKDSKTDGDDANIHLNGTASTLTDTLNLSGAT-TNLG	1497
QY	1250	GNTIKEDKTRAAISIVDVLVSAGFNLOQ-----NGBAVDVFSYTDVTFNPAQGNATPAK	1301
Db	1498	NGCITDNEKKXAAASKVDVLNAGWNVRGKVPASANNQVENIDFVATYDITVDFVSGDKDITTS	1557
QY	1302	VIVD--DTSKTSKVYVDNVDDTTIEVRDKKLGKVTITLTSTG-----TGAN-KFALSNQQA	1354
Db	1558	VTVESKDNGKRTVEV-----KIGATSVIKDHNGKGLTFGKELDANNNGV	1601
QY	1355	T-----GDALVKASDIVAHLNLTSGDITQATKAGSAQNNAGYVDADGNKVIVYDSTD	1405
Db	1602	TVTETDGDDEGNLTYAKAVIDAVNKAGRWKTT-GANGQNDFAIV-ASGTNVTF-----	1655
QY	1406	NKYYQAKNDGTVDKTEKAVOKLVAQAQTPDCTL-AQMNKVSVINQOVNDANKQOG---	1461
Db	1656	-----ADNGGT--TAEVTK-----ANDGSTVKYNVK-----VADGLKLDDGKI	1692
QY	1462	INEDNAFVKLEKASDN-----KTKNAAVTVGDLNVAQPTPLTFAGDTGT-----	1507
Db	1693	VADTTVTLTVADGKVTPAPNNGXGKPEXDASGLAGCLNKLSXT--ATAGEKGTGEYDVPANSA	1750
QY	1508	-TAKLKELTITIKGGOTDNTKLTDMNIGVAGTDGFTVKLAKDLTILNSV-----NAG--	1559
Db	1751	GQEVKAGDAVTEKAG-----DNLKIKOSXKDFYSLKELKDLTISVEFKDANGGTG	1801
QY	1560	--GTKTDDKGVSVFSDSSQQA-----KANTPVLNSANGLDLGKVISNYGKGTK	1604
Db	1802	SESTKTKTKDGLTITPANGAGAAGANTANTISVTKDGISAGKNKATVNVVSGLJK	1853

RESULTS

US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Batenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.

Db 684 -----NCTAVTKGPFVTKGATDGRGVTV-KDATANDAKKVATVK-----DVATA 731
QY 683 KQGTVPF-----GLSQDSGLTIGKSTLNNDGLTVKDT-----NBQIQVGANGIKFT- 728
Db 732 INSAATPVKTENLTTSIDEDNPNDNGK-----DALKAGDPLTFKAGKNLKVKRGKNITF 787
QY 729 -----NVNGSNPGTGIANTARI-----TRDKTGFAGSDCAVDNRPYLD 767
Db 788 DLAKNLEVKAKYSDTLTIGGNTPTGGTTATPKVNIITSTADGLNFA----- 833
QY 768 QDKLQGVNKTITNGINAGKAI--TGLSPTL--PSIADQSSRNIEGNTIQDKDSNAA 823
Db 834 -----KETADASGSKNYLKGIAITLTPESGAKSS-HVDL--NVDATKKSNA 879
QY 824 SINDILNTGPNLNNNNPIDFVSTYDIDVFANGNATATVTHDTANKTSKVVDVNVDDT 883
Db 880 SIEDVLRAGNIQGNNGVYVATYDVTNFTDDSTGTTVT-----V 921
QY 884 TIHLTGDDNKKLGKVTTKLNKTSANGNTAT-----NPNVSSDED-----ALVNAK 930
Db 922 TQADGKGVADYKIGAKTSVTK--DHNGKLTGDKLDANNGATVSEDDGKDTGTGLVTAK 979
QY 931 DIAENLTLAKEIHTTKGTADTALQTFVKKVDENNADANAITVQ-----KNANQVN 986
Db 980 TVIDAVNKSQWRTGEGATETATV-----NAGNAETVTSQTSVNFNGM--AT 1028
QY 987 TLLKKGNG-LNITKDKNGVTVEINTTSLKAG-----KSTLNDGGLSINKNPTGSE 1037
Db 1029 TATVSKDNGNINKYD-----VNVGDLKIGDKKIVADTTLTVTGKVSVPAGAN 1080
QY 1038 QIQ-----VCADGVKFAKVN----- 1053
Db 1081 SVANNKLVNAEGLATA-LNLSWTAKADYADGESEGTQEVKAGDKVTFKAGNKLKV 1139
QY 1054 -----NCVVGAGIDGT-----TRTRDEIGFTGNGS-----LDKSKPH 1087
Db 1140 KQSEKFTYSLQDTLGLTITLGGTANGRNDTGVINKDGLTITLANGAAAGTDSNGN 1199
QY 1088 -----LSKDGINAGKKTINTQSG-----EIAQNSHDVATGGKIYDLKTELENKI----- 1132
Db 1200 TISVTKDGISAGKEITNVKSAKTYKDTQNTADE-TQDKFHAHVKNANEVEFVGKNGA 1258
QY 1133 SSTAKTAQNSLH-----EFSVADEQGNFTVSNPNSSYDTSK 1169
Db 1259 TVSAKTDNNGKHVTTIDVAAKYVGDGLEKDTGDKIKLKVNDTQGNLLTVDATKGSVAK 1318
QY 1170 -----TSDVITPAGENG-----ITTKVNGVVRVGDIDQTKGLT----- 1202
Db 1319 GEFNAVTTDATTAQGINANERGKVVVKGSGATATETDKKKVATVGDVAKAINDAATFVK 1378
QY 1203 ----- 1202
Db 1379 VENDSDATIDSDPTDGDANDALKAXDTLTLKAGKNLKVKRDGKNITFALANDLSVKSATV 1438
QY 1203 TPKLTVCNNNGK-----GIVIDSQNGNT-----ITCLSNLTANVNDKGSVRTTEQ 1249
Db 1439 SDKLSLGTNGKNVNIISDTKGLNFADKSKGTGDDANTHLNGIASTLTDTLNLSGAT-TNLG 1497
QY 1250 GNTIKDEKTRAASIVDVLSAGENLQ-----NGEAVDFVSTYDVTNFDAGNATTA 1301
Db 1498 NGGITDNEKXKXASVKDVLNAGVNRGVRPASANNQVENIDFVATYDVTDFVSGDKDXTTS 1557
QY 1302 VTYD--DTSKTSKVVDVNVDDTTIEVKOKKLGKVTTLTSTG-----TGAN-KFALSNOA 1354
Db 1558 VTVESKDNGKRTVE-----KIGAKTSVKDHNGKLFTEKELKDANNNGV 1601
QY 1355 T-----GDALVKASDIAVHLNLTSGDIOTAKASQANNSAGYVDADGNKVIYDSTD 1405
Db 1602 TVTETGKDEGNGELVTAKAVIDAVNAGVRKTT-GANGQNDFAFV-ASGTNVT- 1655
QY 1406 NKYYOARNDGTVDKTEKAVARDKLVAQAQTPDGTPL-AQMNVKSVINKEQVNDANKKOG--- 1461

Db 1656 -----ADNGNGT---TAEVTK-----ANDGSITVKYNVK-----VADGLKLDGDKI 1692
QY 1462 INEDNAFVKGLEKAASDN-----KTKNAAVTVGDLNNAVAQTPLTFAGDGTG----- 1507
Db 1693 VADTTLTVADGVKVTAPNNGXGKFKXADASGLAGCLNKLST--ATAGKEGTGEVDPANSA 1750
QY 1508 -TAKKLGELTITIGGQDTPNKLNDNNIGVVAGTDFGTFVKLAKDLTNLNSV-----NAG-- 1559
Db 1751 GQEVKAGDKVTFRAG-----DNLKIKQSKDFTYSLKELKDLTSTVEFKDANGGTG 1801
QY 1560 --GTKIKDKKGVSVDFSSGQA-----KANTPVLSSANGLDLGGKVISNVGKGTG 1604
Db 1802 SESTKTKDGLTITPANGACAAGANTANTISVTKDGLISAGNKAVTNVVSGLK 1853
RESULT 6
US-08-409-995-2
; Sequence 2, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RPT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-2

Query Match 6.6%; Score 678.5; DB 1; Length 1098;
Best Local Similarity 23.6%; Pred. No. 1.1e-33;
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;
QY 944 HTTKGTADTA---LQFTTVKKVDENNADANAII-TVGKNANNQVNTL-----TL 990
Db 26 HTKASATVAVATLALLSATVEANNNTPTVTKLKAYGDANFNFTNNSIADEAKQVQEA 85
QY 991 KGBNGLNINIKDNGTGTFTGNTTSLGKAGKSTLNDGGLSINKNPTGSEIQVGADGVKFAK 1050
Db 86 KGLLNLN---EKNASDKLLVEONTAATVGNLRLKGLVLSKNGTRNEKQ---QVKA- 137
QY 1051 VNNNGVVGAGIDGTTRITRDEIGFTGTGNSLDSKPSHLSKDGINAGKKTINTQSGEIAQ 1110
Db 138 -----DEVLFEG-----KGVQVTS----- 152

QY 1111 NSHDAVTGGKIYDLKTELENKISSITAKTAQNSLH--EFSVADEOQGNFTVSNPSSYDTS 1168
Db 153 -----TSNGKHITITFALAKDLG-----VKTA 174
QY 1169 KTSVDITFAG--ENGITTKVNGVVRVIGDQTKGLTTPKLTIV--GNNGKGIVIDS--QNG 1223
Db 175 TVSDTLFIHGGAACAT-----TTPKVNVTSTTDLKFAKDAAGANG 216
QY 1224 QNTI--TGLSNTLANVTNDKGSVRTTEQGNIIKDEK--TRAASIVDVLSAGFNLO---- 1275
Db 217 DTTVHLNGIGSTLDTL--VGSPTHIDGG--DOSTHYTRAASIKDVLNAGWNKIGVKA 271
QY 1276 ---GNGEAVDFSVTYDVNFPADGNATATKVTYDDTSKTSKVYDVNVDDTTIEVKDKKL 1331
Db 272 GSTTQSENVDVHTYDTEFLSADTETTTVDSKENGKRTEYKIGAKTSVIREKDGKL 331
QY 1332 GV-KTTTLTSTGTGANKFALSNOATGALVKASDIVAHLNLTSGDIOTAKGASOANNSAG 1390
Db 332 FTGKANKETHKVDGAN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGONGDFAT 389
QY 1391 YVD-----ADGN--KVIXDSTDKYQAKNDGTVDKTEKVAKDKLVAQAQTPDGTGLAQ 1441
Db 390 VASGTNVTTFASGNTTATVTNGTDG--ITVKYDAKVGDLKLDGDKIAA-----DTTALT 442
QY 1442 MNKSVINKEOVNDANKKOGINEDNAFVKGLEKAASDNK---TKNAAVTVGDNLNAVAQTP 1498
Db 443 VN-----DGKNANPNPKGVADVASTDEKKLVTAKGLVTA--LNSLSWTT 484
QY 1499 LTFAGDTG-----TAKKLGELTITKGGOTDNKLTNDNNGVVGAGTGDGTVKLAKDLT 1551
Db 485 TAAEADGGLDGNASEQEVKAGDKVTFKAG-----NLKVKQEGANFTYSLQDALT 535
QY 1552 NLNSV-----NAGGTIKDKGVFVDSGQA--KANTPVLANSGLDLGGKVISNVGKGT 1603
Db 536 GLTSITLGTGNGAKTEINKDGLTITPANGAGANNANTISVTKDGSAGSGSVKNVYVSGL 595
QY 1604 KDTDAANVQOL-----NEVRNLGLGNAGDN-----ADGNQVNIAD----- 1640
Db 596 KKFGDANFDPLTSSADNLTKONDDAYKGLTNLDEKTDKQTPPVVADNTAATVGDRLGLGW 655
QY 1641 -IKDPSNGSS-----NRTVLKAGTVLGGKGNNDTEKLTATGGIOGVGDKNANGDL 1692
Db 656 VISADKTTGGSTEHQDVNRANEVK---FKSGNGINVGKTVNGRREITFEL---AKGEV 709
QY 1693 --SNWVKTKQDKSKALLAT---YNAAGQTNLYTNPAEAIIDRINEQGRFFHFVNDGNQ 1747
Db 710 VKSNEFTVKETNGKETSIVKLVKGVKYSKEDIDLTTGP-----KLKDN- 753
QY 1748 EPVVOGRNGIDSSAGSKHSVAIGFOAKADGEAAVAIGRQOAGNQSTAIAGDQAATGDQS 1807
Db 754 -----TVAAKYQDK-----GGKVSVDNTTEATITNK 780
QY 1808 IAGTGNVAGKHSGAIGDPSTVKRADNSYSVGNNO-----FTDATQTDVFGVGNNTIV 1861
Db 781 ---GSGVTGQVADAL-----AKSGFELGLADEADAKRAFODKTKALSAGTEIYNA 830
QY 1862 TESNSVALGNSAISAGTHAGTQAK-----KATQSIANAT 1959
Db 831 HDKVFREANGLNTKYSAAVSTESTDANGDKVTTTFVKTDVLPDLTOIYNTDANGKKITKVK 890
QY 1887 -----KSDGTAGTTTT-----AGATGTVKGFAGQTAAG 1914
Db 891 DGQTKWYELNADGTADMTKEYTLGNVDSGKVKVNDGKWKYHAKADDTKTKGSND 950
QY 1915 AVSVG-----ASGAERLIONVAGEVSATSTDAVNGSOLY---KATQSIANAT 1959
Db 951 KVSDEKHVWSLDPNDQSGKGVVDINVANGDISATSDAINGSOLYAVAGVYTNLAGQV 1010
QY 1960 NELDRIHONENKANAGISSAMAMASPMQAYIPGRSMVVTGGIATHNGOGAVAGLSKLSD 2019
Db 1011 NNLSEKVNKVKRADAGTASALASQLPQATMPKGSWVAINGSSYQONGLAIGVSRISD 1070
QY 2020 NGQWFKINGSADTQGHVGAAGAPHF 2047

Db 1071 NGKVIIRLSGTTNSOGKGTGAAGVGYQW 1098

RESULT 7

US-08-685-467-2
; Sequence 2, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 398-3249
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-685-467-2

Query Match 6.6%; Score 678.5; DB 3; Length 1098;
Best Local Similarity 23.6%; Pred. No. 1.1e-33;
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;

QY 944 HTTKGTADTA---LQFTTVKKVDENNADANAI-TVQGNKANNQVNTL-----TL 990
Db 26 HTKASATVAVAVLATLLSATVEANNPTVTKLKAYGDANFNFTNNSIADAQKQVEAY 85
QY 991 KGENCLNKTDKNGTVTEGINTTSLKAGKSTLNDGGLSIKNPCTGSEQIOVGADGVKFAK 1050
Db 86 KGLNLN---EKNASDKLLVEDNTAATVGNLRKLGWLVSSKNGTRNKSQ-----QVKA- 137
QY 1051 VNNGVVAGIDGTTTRIFRDEIGFTGTNGSLDKSPHLSKDGINAGKKITNIQSGEIAQ 1110
Db 138 -----DEVLFEG-----KGVQVTS----- 152
QY 1111 NSHDAVTGGKIYDLKTELENKISSITAKTAQNSLH--EFSVADEOQGNFTVSNPSSYDTS 1168
Db 153 -----TSNGKHITITFALAKDLG-----VKTA 174
QY 1169 KTSVDITFAG--ENGITTKVNGVVRVIGDQTKGLTTPKLTIV--GNNGKGIVIDS--QNG 1223
Db 175 TVSDTLFIHGGAACAT-----TTPKVNVTSTTDLKFAKDAAGANG 216


```
QY 1224 QNTI--TGLSNTLANVNDKGSVRTTEQGNIIKDEK--TRAASIVDVLSAGFNLO----1275
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 217 DTTVHLNGIGSTLTDTL--VGSPTHIDGG---DQSTHVTTRAASIKDVLNAGWNIGVKVA 271
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1276 ----GNGEAVDFVSTYDTNFADGNATTAKVYDDTSKSVVYDVNDDTTIEVKDKKL 1331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 272 GSTGQSENDFVHTYDTVEFLSADTETTTVTVDSEKNGKRTVEKIGAKTSVKEKDGKL 331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1332 GV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIOTAKGASQANNSAG 1390
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 332 FTGKANKETNKVDGAN--ATEDADECKGLVTAKDVIDAVNKTGWRKTTDANCQNGDPAT 389
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1391 YVD-----ADGN---KVYDSTDNKYQAKNDGTVDKTEKAVAKDLVAQAQTPDGTTLAQ 1441
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 390 VASGTVNVTASGNGTTATVTNGTDG--ITVKYDAKVGDLGLKLDGDKIAA-----DTTALT 442
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1442 MNVKSVINKEQVNDANKKOGINEDNAPVKGLEKAASDNK---TKNAAVTVDGLNVAQTP 1498
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 443 VN-----DGKNANPNKGVADVASTDEKRLVTAKGLVTA--LNSLSWTT 484
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1499 LTFAGDTGT-----TAKKLGTELTIKGQTDNTNLTNNIGVAGTDFGFTVKLAKDLT 1551
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 485 TAAEADGGTLDGNASQEVKAGDKVTFKAG-----NLKVKQEGANFTYSLQDALT 535
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1552 NLSNV-----NAGGPKIDDKGVSVFVDDSSGQA--KANTPVLNGLDLGGKVISNVKGT 1603
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 536 GLTSLITLGTGNNGAKTEINKDGLTITPANGAGANNANTISVTIKDGISAGSQSVKNVYSGL 595
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1604 KDTDAANVQOL-----NEVRNLLGLGNAGNDN---ADGNQVNTAD-----1640
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 596 KFGDANFDFLSSADNLTKQNDAYKGLTNLDEKTDKQTPVVAWNTAATVGDRLGLGW 655
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1641 -IKKDPNSGSS-----NRTVIKACTVLGGKGNNDTEKLTGQVGVGDKNAGNDL 1692
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 656 VISADKTTGGSTEYHQPVRNANEVK---FKSGNGINVSQKTVNGRREITFEL---AKGEV 709
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1693 --SNVWVKTKQDKGSKALLAT---YNAAGTQNYLTNNPAAEDRINEQIRFHVNDGNQ 1747
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 710 VKSNEFTVKTNGKETSLSKVGVDKYYSKEDIDLTTGQP-----KLKDN- 753
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1748 EPVVGQRNGIDSSAGSKHSAVGFQAKADGAAVAIGROTQAGNQSIAGDNAQATGDQS 1807
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 754 -----IYAAKYQDK-----GGKVVSYTDNTEATITNK 780
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1808 IATGTGNVAGKHSAGIDPSTVKADNSYSVGNNO-----FTDATQDVFVGVGNNTIV 1861
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 781 ---GSGVVTGNQVADAI-----AKSGFELGLADEADAKRAPDDTKKALSAGTTEIVNA 830
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1862 TESNSVALGNSAISAGTHAGTOAK-----1886
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 831 HDKVRFANGLNTKVSATVESTDANGDKVTTFTVKYTDVDELPLTQIYNTDANGKKITKVVVK 890
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1887 -----KSDGTAGTTTT-----AGATGTVKVGFAGQTAVG 1914
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 891 DGQTKVVELNADGTADMTKEVTLGNVDSGKKVVKDNDGKWKYHAKADGTADTKTGEVSD 950
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1915 AVSVG-----ASGAERIRONVAGEVSATSTDAVNGSOLY----KATQSIANAT 1959
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 951 KVSDEKHVVSLDPNDPQSKGKGVVDNANGDISATSTDAINGSOLYAVAKGVNTLNGQV 1010
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1960 NELDHRHONENKANAGISSAMAMSPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSD 2019
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1011 NNLGKVNKVGKRADAGTASALAAASQLPQATWPKSMVAIAGSSYOCQGLAIGVSRISD 1070
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 2020 NGQWFKINGSADTOGHVGAAGVAGFHF 2047
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1071 NGKVIIRLSGTTNSQKTKGVAAAGVGYOW 1098
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
```

RESULT 8

US-09-377-155-32

; Sequence 32, Application US/09377155

```
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 085064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-32
```

```
Query Match 6.6%; Score 678.5; DB 4; Length 1098;
Best Local Similarity 23.6%; Pred No. 1.1e-33;
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;
```

```
QY 944 HTTKGPDATD---LQFTTVKVDENNADDANAI-TVGOKNANNQVNTL-----TL 990
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 26 HTKASATVAVAVLATLLSATVEANNPTVTKLKYAGDANFNFTNNSADAEEKVQVEAY 85
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 991 KGENGLNIKTNGTTFGINTTSGLKAGKSTLNDGSLTKNPTGSEQIQVGADGVKFAK 1050
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 86 KGLLNIN---EKNASDKLLVEDNTAATVGNLKRKLWLSKNGTRNEK90---QVKHA- 137
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1051 VANNVGVGAGIDGTTTRITRDEIGFTGTNGSLDKSPHLKSKDGINAGGKKTNTIOSGEIAQ 1110
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 138 -----DEVLFEG-----KGVQVTS-----152
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1111 NSHDAVTGKIYDLKTELENKISSTAKTAQNSLH---EFSVADBOGNNTVSNPSYDTS 1168
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 153 -----TSENGKHTTTFALAKDLG-----VKTA 174
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1169 KTSDDVITFAG--ENGITTKVNGVVRVGIDQTKGLTTPKLTV-GNNNGKIVIDS--QNG 1223
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 175 TVSDTLITCGGAAAGAT-----TPPKVNTSTTDDGLKFAKDAAGANG 216
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1224 QNTI--TGLSNTLANVNDKGSVRTTEQGNIIKDEK--TRAASIVDVLSAGFNLO----1275
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 217 DTTVHLNGIGSTLTDTL--VGSPTHIDGG---DQSTHVTTRAASIKDVLNAGWNIGVKVA 271
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1276 ----GNGEAVDFVSTYDTNFADGNATTAKVYDDTSKSVVYDVNDDTTIEVKDKKL 1331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 272 GSTGQSENDFVHTYDTVEFLSADTETTTVTVDSEKNGKRTVEKIGAKTSVKEKDGKL 331
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1332 GV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIOTAKGASQANNSAG 1390
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 332 FTGKANKETNKVDGAN--ATEDADECKGLVTAKDVIDAVNKTGWRKTTDANCQNGDPAT 389
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1391 YVD-----ADGN---KVYDSTDNKYQAKNDGTVDKTEKAVAKDLVAQAQTPDGTTLAQ 1441
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 390 VASGTVNVTASGNGTTATVTNGTDG--ITVKYDAKVGDLGLKLDGDKIAA-----DTTALT 442
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1442 MNVKSVINKEQVNDANKKOGINEDNAPVKGLEKAASDNK---TKNAAVTVDGLNVAQTP 1498
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 443 VN-----DGKNANPNKGVADVASTDEKRLVTAKGLVTA--LNSLSWTT 484
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1499 LTFAGDTGT-----TAKKLGTELTIKGQTDNTNLTNNIGVAGTDFGFTVKLAKDLT 1551
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 485 TAAEADGGTLDGNASQEVKAGDKVTFKAG-----NLKVKQEGANFTYSLQDALT 535
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 1552 NLSNV-----NAGGPKIDDKGVSVFVDDSSGQA--KANTPVLNGLDLGGKVISNVKGT 1603
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 536 GLTSLITLGTGNNGAKTEINKDGLTITPANGAGANNANTISVTIKDGISAGSQSVKNVYSGL 595
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
```


Qy 1902 GTVKGAGQAVGAVSGASCARRIONVAAGEYSATSTDAVNGSQLY---KATQSTANA 1958
Db 538 -----DADKIKVASDGISAGNAKAVNVAAGEISATSTDAINGSQLYAVAKGVNLAGQ 590
Qy 1959 TNELDHRTHONENKANAGISSAMAMSPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLS 2018
Db 591 VNNLEGVKNVKKRADAGTASALAASQLPQATMPGKSWSVTAGSSYQONGLAIGVSRIS 650
Qy 2019 DNGQVFKINGSADTOGHVGAAGAGHFH 2047
Db 651 DNGKVIIRLSGTTNSQKGTGVAAGVGQW 679

RESULT 11
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 5.08; Score 517.5; DB 2; Length 1600;
Best Local Similarity 21.99; Pred. No. 1.5e-23;
Matches 418; Conservative 239; Mismatches 766; Indels 487; Gaps 94;

Qy 1 MNHIYKVFENKATGTFMAVAEYAK--SHSTGGGSCATGQGVSVCTLSFARITAAALVLVIG 58
Db 1 MNKYRLKFKRLNALVAVSRLTRGCDHSTKGEKPVK-----TKVRHLAKPLS 51
Qy 59 ATLSGSAYAOKKDKTKHAIQNGQNPRESGTA--RADGDRAITAIGENAAAA----- 105
Db 52 AILLSLGMASIPQSV-LASGLQGMSVVIHTATMQVDGKNT-TIRNSVNAIINWKQFNIDQ 109

Qy 106 -QGQATAIGASSNKTNYNGSSLDKIG-----TDATQGESIAIGGVKASGDASIAIGSDDL 159
Db 110 NEMEQFLOESSNSAVFNRTVSDQISQLGILDSNGQVEL-----INPNC---ITIGKD-- 159
Qy 160 HLLQDHGNPKPGKTLINDLINGHAVLKEIRSSKDNVDKYRRTTASGHASTAVGAMSTAQ 219
Db 160 -----AIINTNGFTASTLDSINIKARNFTLE---QTKDALAEIV 198
Qy 220 GHFSNAFCTRATAKSAYSGLAVLAATAEGQSTIATIGSDATSSSLGAIALGAGCTRAQLQGS 279
Db 199 NH-----GL-----ITVGKDSVNLIGGKYNKNEGVISVNGS 230
Qy 280 IALGQSVVVTOSDNNRSPAYTPNTQALDPKFOATN-----NTKAGPLSTGSSNSIKRKIINV 335
Db 231 ISLLAGQKITISD-IINPTIVYSIAA--PENEALNGLDIFAKGGINVRAATIRNKGKLS 287
Qy 336 GAGVKNKTDVNV-----AQLEAVV---KWAKERITFGQDDNSDVKIGLDTLITI 383
Db 288 ADSVSKDSKGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGD--KYTLKTGAVIDLSG 345
Qy 384 KGAETNALTDNNIGVVKKEADNSGLKVKLAKTLNLTETVNTTTLNATTTVKVSSSSTTA 443
Db 346 KEGGETYLGDER---GEGKNG---IQAK-----KTTLEKGSTINVSKEKGR 389
Qy 444 ELLSDSLTFTQPTNGSQSTSKTVYGVNGVKTNNNAETAAIGTTRITRDKIGFARDGVD 503
Db 390 AIVMGDIALIDGNINAQGSIAKTG-----GFTVSHGLSIGDVI 433
Qy 504 EKQAPYLDKKOLKVGCSVAITIDN-----GIDAG-----NKKISNLAKSSANDAVTIQOL 553
Db 434 AKEW-LDDPDVDSIETLTSGRNTGENQGYTGGTKESPKGNSISKPLTNS--TLEQI 490
Qy 554 KAAKPTLNAGAGISVTP-----TEISVDAKSGNVTPATNYICGVKTELSNDSGTSDK---F 605
Db 491 -----LRRGSYVNITANNRIYVNSSINLSNGSLTLHKRDGVK---INGDITSNENGL 541
Qy 606 SVKSGS-----TNNSLVTAEHLASLYNEVNRTASALQSFTVKEEDDDANA----- 652
Db 542 TIKAGSWVDVHKNTLGT-----GFLNIV-----AGDSVAFEREGDKARNATDAQITAQ 590
Qy 653 --ITVAKD-----TTNAGAVSILKLGKNGNLVATKK 683
Db 591 GTITVNDKDKQFRFNNVSLNGTGKLGKFIANONNFTKFDGEINI---SGIVTINQTTKK 647
Qy 684 DGTVTFGLSQDSGLTIGKSTLNNDLTVKDTNEQIQVGANGIKFTNV-----NGSNPTGTI 739
Db 648 D-VKYWNASKDSYWNVSSLTLN---TVQ-----KFTFIKFDVSGSN-GQDL 688
Qy 740 ANTARITRDKIGFAGSDGAVDTN-----KPYLDQDKLQ-----VGNVYKIINTGIN 784
Db 689 RSSRR-SPAGVHFNGIGGKTNFNIGANAKALFKLPNAATDPKELPITFANITATG-N 746
Qy 785 AGKKAITGLSPTLPSIA---DQSSRNIELG---WTIQDKDSNAASINDILN---TGNL 835
Db 747 SDSSVMFEDIHANLTSRAAGINMDSINITGGGLDFTSTSHRNSNAFEIKKDLTINATGSNF 806
Qy 836 KNNNNPIDFVSTYD---IVDFANGNATTATVTHTFANKTSKVYDVNVYD---DTTHLTG 889
Db 807 SLKQTKDSFYNEYSKHAINSSHNLTLGNGVNLGGENSSSITGNITNKNANVTLOADT 866
Qy 890 TDDNKKLGVTTKLKNKTSANGN-TATNPNVNSSDBDALVNAKDIAENLNTLAKEIHTTKG 948
Db 867 SNSNTGLKKRLLTLCNISVEGNLSLTGANAN-----IVGNLSIAED-----STFKG 912
Qy 949 TADTALQ---TFTVKKVDENNADDAN---AITVCQKANNQVNTLTLLKGENGLNIKTDK 1002
Db 913 EASONLNTIGTFT-----NNGTANINIKGVYKLGDN-----NKGGLNITINA 955
Qy 1003 NGTVTFGLNTTSGLKAGKSTLNDGGLSLKNPTGSGQIOVGADGVKFAKVNNGVVGAGID 1062
Db 956 SGTOKTIIN-----GNITNEKGLNINIKAKADEIQIGN-----ISOK-----E 995

QY	1063	GTTTRITRDEIGFTGNGSLDKSPHLSDKGINAGKKITNIQSGEIAQNSHDAVTGGKYY	1122
Db	996	GNLTISDDXVNIIT-----NQTIKAGVEGGRSDSSEAEANANLIQTQKELKLAG---	1043
QY	1123	DKLTELENISSAKTAQNSLHEFSVADEQGNFTVSNPYSSYDTSKTSVDITFAGEN--	1180
Db	1044	DUNISGFNAEITKANGS-----DLTIGHASGNDADAK--VTFDKVKDKLIST-DGHNVT	1096
QY	1181	-GITTVMKGVVRVIGDQTKGLT--TPKLTVGNNGKGVIDSGONGQNTITGLSNTLANV	1237
Db	1097	LNSEVKTSNGSSNAGNDNSTGLTISAKDVTVMNNVTSHTKINISAAAAGNVTTEGTTINA	1156
QY	1238	TNDKGSVRTEBOGNIIDKEDKTRAASIYDVLSAGFNOLOGCEAVDFYSTYDTVNFADGNA	1297
Db	1157	TT--GSVEVTAQNGTIK-----GN-----ITSQNVTVTATENLV	1188
QY	1298	TTAKVYDDTSTSKVYDVNVDDTTIEVKDKLGVKTTT---LTSTGTGANKFALSNO	1353
Db	1189	TTENAVINATSGT-----VNISTKTGDIKG--GIESTSGNVNITASG---NT'LKYSN-	1235
QY	1354	ATGDALVKASDIVAHLNLTLSGDIQTAKG-----ASQANNSAGYVDADGNKVIYDSTDNKYY	1409
Db	1236	ITGQDVTVTADAGALTTTAGSTISATTGNANITTKTGIDNGKVESSGSVTLATGATLA	1295
QY	1410	QAKNDG-TYDKTKVEAKDKLVAQ-AQTDPGTLAOMNVKSVLNKEQVNDANKKQGINEDNA	1467
Db	1296	VGNISGNTVITRAD--SGKLTSTVGSTINGT-----NSVTTSQSGD-----	1335
QY	1468	FVKGLEKAASDNKTKNAAVTVGDLNNAVAQTPLTFAGDTGTTAKKLGETLTIKGGQDTNKK	1527
Db	1336	-----IBCTISGN-TVMVTASTGD-----LITGNSAKVBAKNGAATLTAESG-----K	1377
QY	1528	LTDNNIGVAGTDGFTVKLAKDLTNLNSVYNAAGTKIDDKGVSEVDSSGOAKANTPVL SAN	1587
Db	1378	LTTQTGSSITSSNGQTTLAKDSSTAGNINAANVLTNTGTLTITGDSKINNTSGELTIN	1437
QY	1588	GLDL-----GGKVISNV-----GKG-----TKDPTDAANYO-OLNEVRNLLGLGNAGNDNA-	1631
Db	1438	AKADKLDAAGSGDRTVVNAATNATNAGSGNVTAKTSSSVNITGDLNTINGLNIISENGRVTNR	1497
QY	1632	-DGNOVNIADIKKDPNSGSSSNRTVTKAGTVL-----GGKGNNDTEKLATGGIO-VGVDPKG	1686
Db	1498	LRGKIEDVKYIQ-----PGVASVEEVLKAKVLEKVKDLSDERETLAKLGVSAVRFVEPN	1553
QY	1687	NANGDLSNWMVKTKQDGSKALLATVYNAAGOTNYLTNNPAEADRINQQ	1736
Db	1554	NA-----IVNTONEFTTKPSOVTISEGKACFSGSGNGARVCTNVADGG	1597

RESULT 12

US-08-617-697-9

; Sequence 9, Application US/08617697

; Patent No. 5977336

: GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

: TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: 5977336-Typeable Haemophilus

FILE OF INVENTION:	11
NUMBER OF SEQUENCES:	11
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1

NUMBER OF SEQUENCES: 1.
CORRESPONDENCE ADDRESS:

ADDRESSEE: shoemaker and Mattare. Ltd.

ADDRESSEE: SIOEMAKER and MALLORY, LTD.
STREET: 2001 Jefferson Davis Hwy. 1203 Crystal Plaza

STREET: 2001 JE
STREET: Bldg 1

STREET: Bldg. I
CITY: Arlington

CITY: ARLINGTON
STATE: VIRGINIA

STATE: Virginia
COUNTRY: U S A

COUNTRY: U.S.A.
ZIP: 33303-0786

; ZIP: 22202-0286
; COMPUTER READABLE FORM:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS

```

; SOFTWARE: PatentIn Release #1.0,
; CURRENT FOOT LOCKING DATE:

```


Db 606 VLSNLGKLSFTDSREDRGRRTKG--NISNKFCTGLNISCTVDSMKAPKVSFWRDKG 663
Qy 1004 CT---VTEGINTTSLKAGKSTLNDG---GLSIKNTPTSGEIQVGADGVKFAKVNNGV 1056
Db 664 RTYNNVT--TLNVTSGSKFNLSDTSGSGTSPISRN-----AELNGITFNKATFN-- 712
Qy 1057 VGAGIDGTRITREIGF--TGTNLSLKSPLSKDG-----INAGGKKTNIOS--GBIA 1109
Db 713 IAQSTANFSIKASIMPEKSNANTALFEDISVSGGSVNFKLASS---SNIQTGPII 769
Qy 1110 QNSHDVATGGKIYDLKTELENKISSTAKTQNSLHEFSVADEQGNFTVSNPYSSYDTSK 1169
Db 770 KSONFWSGGTLLNLAB-----GSTETA-----FSIENDLNUNATGNN----- 808
Qy 1170 TSDWITFAGENGITTKVNGVYRIGIDQTKGLTTPKLTPTGVNNG---KGIVIDSQNGQN 1225
Db 809 ----ITIRQVEGTSRVNKGVAANKNITFG---GNITFGSOKATTEIKGNVTINKNTNA 861
Qy 1226 TITGLSNTLANVTNDKGSVRTEQGNIIKDEKTRAASIVD-----LSAFENLQG--NGE 1279
Db 862 TLRG-----ANFAENKSLNTA--GNVINGNLTAGSIINAGNLTVSKGANLQAITNY 914
Qy 1280 AVDFVSTVD-----TVNFAFGDNATTAKVTYDDTSTKSVVYDVNVDFTIEVKDKLGVK 1334
Db 915 TPNVAGSFDNGASNISARGA-----KPKDINNTSSLNITNSDITYRI----- 961
Qy 1335 TTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIQAKGASQANNSAGYVDA 1394
Db 962 -----IKGNISNKSGLDNIIDKSDAEIOIG---- 987
Qy 1395 DGNKVIYDSTDNKYIOAKNDGTVDKTEKVAQDKLVAQAQTPDGLTQAQNVKSVINKEOVN 1454
Db 988 -----GNISQKEGHLTITSSDRVN 1005
Qy 1455 DANKKOGINEDNAFVKGLEKAAD--NKTKNAAVTVDGLNVAQOT--PLTFAGDTGTAKKL 1512
Db 1006 -----ITNQITTKAGVEGRSSSEANLTI-----QTKELKLAGDLNLISGNK 1051
Qy 1513 GETLTKGGOTDNKLTNNIGVAGTGDFTVKLAKDLTLNLSVSNAGGTKIDDKGVSEVD 1572
Db 1052 AE--ITAKNG-----SDLTIG-----NASGKNADAKKVF--D 1080
Qy 1573 SSGQAKANTPVLNGLDLGKVISNCKGKTGTDAAVVOQLNEVRNLLGLGNAGNDNAD 1632
Db 1081 KVKDSKIS-----TDGHNVTLSNVEYKTSNGSSNAGNDNST 1115
Qy 1633 GNOVNIADIKDPNPGSSSNRTV---IKAGTVLGKGNNDTEKLTATGIGVGVKDGKGNAN 1689
Db 1116 GLTISAKDV--TVNNVTSHKTIINISAAAGNVTYKEG--TINATGGSVEV--TAONGTIK 1170
Qy 1690 GDLSNVWYKQKDGSKKALLATYNAAGOTNYLTNNPABEIDRINEQGIIRFFHVNDGNOEP 1749
Db 1171 GNITS-----QNVTVTATENLVTTENAVINATSGTVNISTKTG----- 1208
Qy 1750 VVOGRNGIDSASCKHSAVGAFOAKDEAAVAGRTQAGNQSIAGDNAAQ--TGDSOI 1808
Db 1209 ---DITKGGEISTSGNNITASGNTLKVSNITGQDVTVTADAGALTTAGSTISATGNAI 1266
Qy 1809 AIGTGNVYAGK---HSGAI-----GDPSTVAKADN-----SYSVGNNN 1842
Db 1267 TTKTGD--TKNGVSSGSGVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNS 1325
Qy 1843 QFTDATQDVFVG--GNNITVYES--NSVALGSNSAISAGTHAGTQAKKSDGTAGTTTAT 1899
Db 1326 VTTSSQSGDIEGTISGNTVNVNTASGDLTIGNSAKVEAKNGAATLTAES--CKLTJTQTGSS 1384
Qy 1900 ATGTVKGFAGQATVAGSVAGSAGERRQNVAAAGVYSATST-----DAVNGSQL 1948
Db 1385 ITSS-----NGQTTTAKDSSJAG-----NINAANVTLLTGTTLTTGDSKINATSGTLT 1434
Qy 1949 YKATOSIANATNELDHRTHQENKANAGISSAMAMASPMQVAYIPGRSMVTGGIATHNGQG 2008

Db 1435 INAKDAKLDGAASGDRTVVNNATNAGSGNVTAKTSSSV-----NITGDLNTIN--- 1482
Qy 2009 AVAVGSLKSLDNGQWVFKING 2029
Db 1483 ---GLNIISENGRNTVRLRG 1499
RESULT 13
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-728-470-10
Query Match 4.9%; Score 501.5; DB 2; Length 1529;
Best Local Similarity 21.3%; Pred. No. 1.3e-22;
Matches 373; Conservative 238; Mismatches 670; Indels 471; Gaps 85;
Qy 425 TTLNATTVKVSSSTTAEALLSDLTFTQPNVTGSGSTKTVYGVNGVKFTNNAETAAI 484
Db 3 SVVHGTTATQVDGNKTTIRNSVNAIINNKQFIDQNE-----QFLQESSNSAVF 53
Qy 485 GTTRITRDKIGFARGDDYDEK--QAPYLDKKQLKQVSVAITDNGIDAGNKKISLNLAGSS 543
Db 54 --NRVTSQISQLK--GILDSNGQVFLPNPNTIGTKDAIINTNGFTASTLDISN--ENIK 108
Qy 544 ANDAVTIEQL--KAAKPTLNAG-----AGISVTPTEISVDAKS-- 579
Db 109 ARN--FTLEQTRDKALAEIVNHLITVGVKGSVNLIGGKVKNEGVISVNGGSIILLAGQKI 167

QY 580 --GNVTAPTNYGVKTELSNDSGTSKFSVKSGTNNLSLVTACHLASYLEVNRVARTADSAL 637
Db 168 TISDIINPTITYSIAAPENAINGLDIFA-KGGNIN-----VRAATIRNKGKLSADSV- 219
QY 638 QSFTVKEEDDDANATVAKD-TTKNAGAVSILKLKGNKGLTVAT----- 681
Db 220 -----SKDKSNIIVLSAKEGEAEGVISAQOQAKGKGLMITGDKVTLTKGAVIDLIS 272
QY 682 KKDGTWTFGLSQD-----SGLTIGKSTLNNDLTVKDTNEQIQVANGIKFTNV----- 730
Db 273 GKEGGETY-LGGDERGEGNGIQLAKKTTLEKGSTINVSCK--EKGGRAIVWMDIALIDG 329
QY 731 NGSNPGTGANTARIYRDKIGFA-----GSDGAVDTNKPYLQDQKLQGVNWKI--T 779
Db 330 NINAQSDIAKTG-----GFVETSGHDLISIGDDVIVDAKEWLLDPDDVSIETLTSGRN 382
QY 780 NTGINAGGRAITGLSPTLPSIADQSSKNIELGNTIQDKDSNAASINDILNTG--ENLKN 837
Db 383 NTGENOG--YTTG-----DGHKSPKGSISKPLITN-STLEQILRQSGYVNITA 429
QY 838 NNPIDFVSTYDIVDPANGNATTATVHTDANKTSKVWYDVNVDDTTIHLTGDDNKKLG 897
Db 430 NNR-----IYVNSSINLSGSLTL-----HTKRDGVKINGDIT-----SNENGLT 470
QY 898 VKTTKLKNTSANGTATNP-NVNSDEEDALVNAKDIAENLNTLAKEIHTTKGTADTALQT 956
Db 471 IRAGSVWDVHKNTLTGTLNIVAGDSVAFEREGDKARN-----ATDAQITAQGT 520
QY 957 FTWKVKYDEN--NNADDANAITVGOKNANNVNTLTGKENGILNKIKTNGVTTFPOINT 1013
Db 521 ITVKKDKQFRFNV-SINGTKGLKFIANO-NNFTHKFDGELNI-----SGVITINQTKR 574
QY 1014 SGLK-----ACK-STLNDGGLSIKN-----PTGSEQIQVGADGVKFAKVNNGVGA 1059
Db 575 KVKVYNASKDSYWNVSSLTNTVOKFTFKFVDSGSSNSQDLRSSRRSPAGVHFNIGICK 634
QY 1060 GIDCTTRITRDEIGFTGWSLDSKPHLSKDGINAGGKIITNIQSGEIAQNSHDAVTGG 1119
Db 635 -----TNFNIGANAK-AFLKLPNAATD-----PKK-----ELPITFANITATGN 674
QY 1120 K-----IYDLKTELENKISSTAKTAQNLHEFVSADQGNFTVSNPYSSYDTSKTSVDIT 1175
Db 675 SDSSVMFDIHANL-----TSRAAGINMDSINITGCLDFSIHSHRNSNAFELKDLTLIN 728
QY 1176 FAGENGITTKVKNGVVRVIDOTKGLTTPKLTGVNNGNGKVIDSON-----QONTITGLS 1231
Db 729 ATGSN-----FSLKQTKD-----SFYNEYSKHAINSSHNLTLGNGVTLGGE 770
QY 1232 NTLANVTNDKGSVRTTEQNIILKDED-----KTRAASIVDVLVSAGFNLOGNGEAVDF 1283
Db 771 NSSSIT-----GNINITNKANVTLOADTSNSNTGLKRTLTGLNISVEG-NLSLTGANANI 826
QY 1284 VSTYDTVNFADGNATKAVTYDDTSKTSKVWYDVNVDDTTIEVKOKKLGKVTTLTSTGT 1343
Db 827 V-----GNLSTA-----EDSTFK-----CEASDNLNIOTGT 851
QY 1344 GANKFALSNOATGDALVKAASDVAHLNLTLSGDIOTAKGASQANNSAGYVDADGNKVYDS 1403
Db 852 -----FTNNGTANINIKGVV-----KLQGDINKNGGLNITNASG-----TQKTLIEN 896
QY 1404 TDNKYYQAKNDGTVDKTEKAVDKLVAQAQTPDGTLAQNNVKSVINKEOVNDANKKOGIN 1463
Db 897 ITNE-----KGLDNINKIKADAIEQI-----GGNISQKEGMLTISSDKVN-----IT 938
QY 1464 EDNAPVKGLEKAASD-NKTKNAAVTVGDLNVAQAQ--PLTJFAGDTGTAKKLGELTLTKGG 1521
Db 939 NQITIKAGVEGGRSDSSEANANLTI-----QTKELKLAGDLNLSGNPKAE-ITAKNG 990
QY 1522 QTDNKLNTNNNGVAGTGDFTVKLAKDLTNLSNVNAGTKIDDKGVSPVDSGQAKANT 1581
Db 991 -----SDLTIG-----NASGNDADAKKVTF-DRVKDSKIS- 1019

QY 1582 PVLISANGLDLGGKVISNVGKGTDKDTPDAANNVQOLNEVRNLLGLCNAGNDNADGNOVNITADI 1641
Db 1020 -----TDCHNVTLNSEVKTSNGSSNAGNDNSTGLTISAKDV 1055
QY 1642 KKDPSNGSSSNRTV--IKAGTVLGGKGNNDTEKLTAGGQVGVDRKDGNGANGDLSNVVVK 1698
Db 1056 --TVNNVNTSHKTIINISAAAGNVTTKEGT--TINATVGSVEV-TAONGTIKGNITS----- 1106
QY 1699 TQKDGSKKALLATYNAGQTNLYLTNNPAEADIRNEQGIREFHVNDSGNOEPVQVQGRNGID 1758
Db 1107 -----QNVTVTATENLVTTENAVINATSGTVNLSITKG-----DIKGGIE 1146
QY 1759 SSASGKHSVAIGFOAKADGEAAVAICRQIQAGNOSIAIGNAQA-TGDSQSIATGTGNVVA 1817
Db 1147 STSGNVIITASGNITLKVSNITGQDVTVTADAGALTITTAGSTISATTGNANITTKGD-IN 1205
QY 1818 GK---HSGAI-----GDPSTVKADN-----SYSVGNNOQTDTATQTD 1851
Db 1206 GKVESGSGSVTLVATGATLAVGNISGNVTITADSKLSTVSGSTINGNSVTTSSQSGD 1265
QY 1852 VFGV--GNNTVITES-NSVALGSNSAISAGTHAGTQAKKSDGTAGTITTAGATGTVKGFA 1908
Db 1266 IEGTISGNVTNYTASTGDLTIGNSAKVEAKNGAATLTAES-GKLTTQTGSSITSS-----N 1320
QY 1909 GQFAGVAGSVGASGABERRIQNVAAGEVSATST-----DAVNGSOLYKATOSIAN 1957
Db 1321 GOTTITAKDSSIAG-----NINAANVTLTNTTGLTTTGDSKINATSGTLTINAKDAKLD 1374
QY 1958 ATNELDHRITQENKENAKAGISSAMAMASMPQAYIPGRSVMTGGIATHNGOGAVAVGLSKL 2017
Db 1375 GAASGDRTVVNATNAGSGCNVTAKTSSSV-----NITGDLNTIN-----GLNII 1418
QY 2018 SDNGQWVFEXING 2029
Db 1419 SENGRTNVTBLRG 1430
RESULT 14
US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:

TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 4.4%; Score 457; DB 2; Length 1338;
Best Local Similarity 21.1%; Pred. No. 6e-20;
Matches 319; Conservative 213; Mismatches 557; Indels 426; Gaps 67;
QY 653 ITYAKDTTKNAGAVSILKL-KKNGLTVATKKDGTFTFG---LSQDSGLTIGKSTLNNDG 708
DB 13 INVRAATIRNKKLSADSVSKSGNIVLSAKEGEAEIGGVISAQNOQAKGKLMITGDK 72
QY 709 LTVKDTNEQIQVQANGIKFTNVGSPG---TGIAANTARITRDKIGFAGSDGAVDTNKP 765
DB 73 VTLK-TGAVIDLSEKGEETYLGGDERGEGKNGIQLAKKTLEK---GS-TINVSGRE 125
QY 766 LDDKQLQGVNKVTNTGINAGGKAI--TG-----LSPTLPS 799
DB 126 KGGRAIWMGDIALIDGNAQAQDIARTGFGFVETSGHYLSIDDNAIVKTKEWLLDPENVT 185
QY 800 IADQSSRNIELGNTIQDKOKSNAASINDILNTGFLNKNNNPDIFFVSTYDVIDVDFANGNAT 859
DB 186 IEAPSASRVELG---ADRNHSAEVIK-----VTLKKNNTSLTTTNTTISNLLK---- 232
QY 860 TATVTHDTAKNTSKVYDVNVDDTTIHLTGTDNKKLGVKTKLNKTSANGNTATN---- 915
DB 233 SAHVNVNITARRKLTVNSSISIERGSHLLHSEGGGGVQIDK-DITSEGNLTIIYSGW 291
QY 916 -----FNVNSDEEDALVNAKDIAENLNLAKIEHTTKGTADTALQTFVVK-- 960
DB 292 VDVHKNTILGSGFLNITTEKGFADFKSGRNNLTITAQGTITSGNSGNFRFNNVSLNSL 351

QY 961 --KYDENNNADANATVCGKNANNQVN-TLTLKNGENGLNIKT-----DKNGT---V 1006
DB 352 GKKLSFTSDREDRGRRTKG--NISNKFDTGLNISGTVDISMKAPKVSFWRDKGRTYWY 409
QY 1007 TFCINTTSGLKAGKSTLNDG---GLSIKNPTGSEQIOVCADGVKFAKVNNGVVCAGID 1062
DB 410 T-TLVNTSGSKFNLSIDSTGSGSTGPSIRN-----AELNGITFNKATFN--IAQGST 458
QY 1063 GTTTRITRDEIGF-TGTNGSLDKSKPHLSKDG-----INAGGKKITNIQS-GEIAQSHDA 1115
DB 459 ANFSIKASIMPKSNANYALEFNEDISVSGGGSVNFKNASS---SNIQTPGVLIKSONFN 515
QY 1116 VTGGKIYDLKTELENKISSTAKTAQNSLHEFVSADQGNFTVSNPYSSYDTSKTSDDVT 1175
DB 516 VSGGSTLNLKAE-----GSTETA-----FSIENDNLNLTATGNN-----IT 550
QY 1176 FAGENGITTKVANKGVVRVGDIDQTKGLTTPKLTJVGNNG---KGIVIDSONCONTTIGLS 1231
DB 551 IQVEGTDSRVNKGVAACKNITFKG---GNITFGSKATTEIKGNVTINKNTNATLURG-- 605
QY 1232 NTLANVTNDKGSVRTTEQGNIIKDEKTRAASIVDY-----LSAGFNLOG-NGEAYDVS 1285
DB 606 ---ANFAENKSPNLNIA--GNVINNGNLTTAGSIINAGNLTVSKGANLQAITNVTENVAG 660
QY 1286 TYD-----TVNFDAGNATTAKVTYDDTSKTSKVYDVNVDDTTIEVKDKLGVKTTLS 1340
DB 661 SFDNNGASNISARGGA-----KFKDINNTSLNITNTSDTYYRTI----- 701
QY 1341 TGTGANKFALSNOATGDALVKASDIVAHLNTLSGDTQTAKGASOANNSAGYVDADGNKVI 1400
DB 702 -----IKGNISNKSGLDNIIDKSDAEIQIG----- 727
QY 1401 YDSTDNKKYQAKNDGTVDKTEKVAQKDLVAQAQTPDGTIAQNMVKSVINKEOVNDANKKQ 1460
DB 728 -----GNISQKBEGLTISSDKVN----- 745
QY 1461 GINEDNAFVKGLEKAASD-NKTKNAAVTYVDLNAVAQT-PLTFAGDTGTAKKLGETLTI 1518
DB 746 -ITNQITIRAGVEGGRSDSEAEANLTI-----QTKELKLAGDLNLTSGFKNAB-ITA 796
QY 1519 KGGQTDNKLTDNNIGVWAGTDGFTVKLAKDLTNLNSVAGGTFKIDDKGVSVFVDSGQAK 1578
DB 797 KNG-----SDLTIG-----NASGGNADAKKVTF-DKYKDSK 826
QY 1579 ANTPVLSANGLDGGKVISNVGKGTDTDAANVQQLNEVRNLLGLNAGNDADGNQVNI 1638
DB 827 IS-----TDGHNVTNLSEVKTSNGSSNAGNDNSTGLTISA 861
QY 1639 ADIKKDPNSGSSSNRTV---IKAGTVLGGKGNNDTEKLTGGIQGVGDKNANGDLSNV 1695
DB 862 KDY--TVNNVTSHKNTINISAAAGNVTTEGT--TINATGSEV-TAQNGTIKGNITS- 915
QY 1696 WVKTKDGSKKALLATYNAAGQNYLTNNPAEAIDRINEQGRFFHVNDCNQBPVVQGRN 1755
DB 916 -----QNVVTATENLVTTENAVINATSGTVNISTKTG-----DIK 952
QY 1756 GIDSSAGSKHSAIGFOAKADGEAAVAIGRQOAGNQSTAIGDNAQA-TCDQSIALGTGN 1814
DB 953 GIESTSGNVTITASGNTLVSNITGDQVTVTADAGALTITAGTSTIATTCNANITTKTD 1012
QY 1815 VWAGK---HSGAT-----GDPSTVKADN-----SYSVGNNOQTDDAT 1848
DB 1013 -INGKVESGSGSVTLVATGATLAVGNISGNTVITADSGKLTSTVSGTNGTNSVTTSSQ 1071
QY 1849 QTDVFGV--GNNITVTES-NSVALGNSAISACTHAGTQAKKSDGTAGTTTATAGATGVK 1905
DB 1072 SGDIETISGNTVNTASTGDLTIGNSAKVEAKNGAATLTAES-GKLTQTGTGSSITSS-- 1128
QY 1906 GFAGQTAQVAVSVCASCAERIONVAAGEVSATST-----DAVNGSOLYKATQS 1954
DB 1129 --NGQTTLTAQSSSIAG-----NINAANVTLTNTTGTTLTTGDSKINATSGTLINAKDA 1180
QY 1955 IANATNELDHRIHONENKANAGISSAMAMPMQAVIPGRSMVTGGIATHNGGAVAVGL 2014

Db 1181 KLDGAASGDRTVWATNASSGNVTAKTSSV-----NITGDLNTIN-----GL 1224
Qy 2015 SKLSDNGOWFEKING 2029
Db 1225 NIISENGENTVRLRG 1239

Search completed: September 13, 2001, 12:37:49
Job time: 48185 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: September 13, 2001, 03:35:34 ; Search time 76.31 Seconds
(without alignments)
2043.367 Million cell updates/sec
Title: US-09-361-619-7
Perfect score: 10303
Sequence: 1 MNHIYKVFINKATGTFMAVA.....NGSADTQGHVGAAGAGHFH 2047
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1143	11.1	2059	D82671	surface protein XF
2	995.5	9.7	1588	A86036	probable adhesin Z
3	778	7.6	1190	A82615	surface protein XF
4	641	6.2	5188	B85547	probable RTX faml
5	627	6.1	2249	A41477	190K surface anti
6	582	5.6	2340	B71704	cell surface prote
7	570.5	5.5	4919	T31105	hypothetical prote
8	521	5.1	2660	B85822	probable invasin Z
9	518	5.0	3029	S76109	hypothetical prote
10	513.5	5.0	2020	C48399	A5C-type transport
11	509.5	4.9	5627	C83339	hypothetical prote
12	506	4.9	4152	T31102	filamentous hemag
13	494.5	4.8	2232	T34434	hypothetical prote
14	491.5	4.8	13288	T03099	mucin, submaxillar
15	479	4.6	2468	A83412	hypothetical prote
16	478.5	4.6	2383	D84962	probable membrane
17	477.5	4.6	2514	F81045	hemagglutinin/hemo
18	471	4.6	1577	A35140	hemolysin A precu
19	466	4.5	1975	B81192	hemagglutinin/hemo
20	462	4.5	2893	A64556	toxin-like outer m
21	457.5	4.4	1645	JN0896	crystalline surfac
22	456.5	4.4	1029	T30852	outer membrane pro
23	456	4.4	1643	D71630	outer membrane pro
24	452	4.4	2055	T31110	extracellular matr
25	452	4.4	2703	D81193	hemagglutinin/hemo
26	451.5	4.4	1902	T06997	lactocepin (EC 3.4
27	448	4.3	3591	S21010	filamentous hemag
28	447.5	4.3	1902	B44858	lactocepin (EC 3.4
29	447.5	4.3	1902	B45784	lactocepin (EC 3.4

hemagglutinin/hemo
outer membrane pro
toxin-like outer m
lactocepin (EC 3.4
hypothetical prote
RTX toxin RTXA VCL
high-molecular-we
probable hemagglut
toxin-like outer m
hemagglutinin-like
hypothetical prote
high-molecular-we
hypothetical prote
hemagglutinin-like
hemagglutinin-like
hemolysin A - Serr

30 446.5 4.3 1995 2 G81044
31 443 4.3 1651 2 JC1340
32 443 4.3 2529 2 B64635
33 441.5 4.3 1962 2 A32634
34 440 4.3 2015 2 B81989
35 437 4.2 4558 2 C82199
36 433.5 4.2 1536 2 A43855
37 433.5 4.2 3535 2 E83641
38 425.5 4.1 2902 2 C71953
39 425 4.1 3282 2 E82750
40 421 4.1 1417 2 A83080
41 419 4.1 1477 2 B43855
42 419 4.1 2154 2 F83068
43 418 4.1 3442 2 B82589
44 418 4.1 3455 2 B82519
45 417 4.0 1608 2 A28182

RESULT 1
D82671
surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82671
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82671
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2059 <SIM>
A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1529

Query Match	11.1%	Score 1143;	DB 2;	Length 2059;
Best Local Similarity	24.0%	Pred. No. 1.4e-36;		
Matches	547;	Conservative 342;	Mismatches 766;	Indels 626; Gaps 104;
Qy	86	SGTA--KADGDRAIAGNANNAQGGQAIAGSSNKT--VNGSSLDKIGTDATGQSSIAIG	141	
Db	86	SGTAAEQCCASNLPLGGSLYVNSGVGVNDVLNKTYSIRMGSVITMTVA-GTNAIATG	144	
Qy	142	GDVKASGDASIAIGSDLLHLLDQHGKPKGTGLNDLINGHVLKELRSSKNDVKYR	201	
Db	145	SAQSSAADAL-----KASL-----A	159	
Qy	202	TATAGHASTAVGAMSYAOCHFSNAPGTRATKAYSIAVLAAATAGOSTIAIGSTSS	261	
Db	160	TKASARAIAIGAKASADGVDTVAIGSGATA-----GTGASSIAIGLNASAV	206	
Qy	262	SIGATAGAGTQAQL-QGSIALQGGSVVTQSDNNRSPAYTPNTQALDPKFQATNNTKAGP	320	

probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A86036
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1588 <STO>
A:Cross-references: GB:AE005174; NID:q12518349; PIDN:AAGS9749.1; GSPDB:GN00145; UWGP:250
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z5029

Query Match 9.78; Score 995.5; DB 2; Length 1588;
Best Local Similarity 23.0%; Pred. No. 4.8e-31;
Matches 505; Conservative 267; Mismatches 665; Indels 761; Gaps 95;
Qy 1 MNHLYKVFENKATGEMAVAEYAKSHSTGGSCATGQVGSVCTLSFARIALALVIVICAT 60
Db 1 MNKIFKVIWNPATGNYTSETAKSRGKKGSK-----LILSALVAGGM 45
Qy 61 LS--GSAYAKKDKTHIAIEQONPRRGTAKADGDRARIAIGENANAGGQAIAGSSN 117
Db 46 LSSFGALANAGNCGVDYG-----SGSA--GDGWAIGKCAK-----N 84
Qy 118 KTVNGS-SLDKIGTD--ATGQESIAIGDVKASGDASTAIGSDLLHLDDQHGNPKHPKGT 174
Db 85 TFMNTSGSTAVGYDAIAEGYISGAIGSKTHAIGGASMAFGVSAT-----129
Qy 175 LINDLINCHAVLKEIRSSKNDVRYRRTTASGHASTAVGAMSYAOGHESNAGFTRATAKS 234
Db 130 -----SEGDRSIALGASSYSLGQYSMALGRYSKALG 160
Qy 235 AYLAVGLAATAEGOSTIAIGSDATSSSLGAIALGAGTGAOLQGSIALGQSVVVTQSDNN 294
Db 161 KLSIAMGSSKAEGANALATGNATKATEIMSLALGDTANASKAYSMALGASSVASEB--217
Qy 295 SRPAPYPMQALDKFQATNTKAGPLSIGNSIKRKLIIN--VGAGVKNKTDVAVNAOLE 351
Db 218 -----NAIAIGAEATEAENATA-----IGNNA-KAKGTNSWAMFG-SLADKVN-----259
Qy 352 AVVKWAKERRITFOGDDNSTDKIGLDNLTIKGAETNALTDDNIGVVKADNSGLKVK 411
Db 260 -----TIALNGSOALADNAIAT-----GOGNK--282
Qy 412 LAKTLNNLLEVTNTTANATTVKVGSSSTTAELLSLSLTFTQPTNGSQSTSKTVGVNG 471
Db 283 -ADGVDAIALGNGSQSRLNTIALGTASNAICD--KSLAL-----GNSNAN--GINS 330
Qy 472 VKFTNNAETTAAGTTRIRDKIGFARGDGVDEKOAPYLDKQKLVGSAITIDNGIDAG 531
Db 331 V-----ALGADSIAT-----DLDDTVSVG 348
Qy 532 N-----KITSNLAKG--SSANDAVTIEOLKAAPT-----LNAGAGISVTPTEISVDAKSG 580
Db 349 NSSLKRRKLVNKNKAIGKSDSDYDAINGSOLYAISDSVAKRLGGGA-----AVDVIDG 399
Qy 581 NVTAPTYNIGVKVTKELNSDGTSDKFSVKGSGTNNLSVTAELHSLAYLNEVNRNRTADSLQSF 640
Db 400 TVTAPTYNL-----KNGSKNNVGA-----L 420
Qy 641 TVKEDDDANAITVAKDPTTNAGAVSILKLGKNGLTVAATKK-----DGTVTFTGLSQD 694
Db 421 AVLDEN-----TLQWDOTKG-----KYSAAHGTSSTPASVITDVAAGTIS-----460
Qy 695 SGLTICKSTLNDGLTVKTDNEQIOVGANGIKFTNVNCSNPCTGIA-----NTARITR--D 748
Db 1328 -----LDNRVTNLDISRV--NIENGI-----GDIV-----T 1351

Db 461 ---ASSKDAVN--GSQKATNDNDVEANTANIA---TNTSNIATNTANTATNTNTNUTD 512
Qy 749 KIGFAGSDGAV--DTNKPYL---DQDKLVQGVNKIITNGINAGGKAITGLSPTLSIADQ 803
Db 513 SVGDLQADALLWNETKKAFSAHGDITS---KITNV-----546
Qy 804 SSRNIELGNTIQDKDKSNAASINDILNTGFLNKNNNPIDFVSTYDIYDFANGNATTATV 863
Db 547 -----KDALITADSTDAVN--GSQKLTNDAV-----ATN 574
Qy 864 THDTANKTSKVYVD--VNVDDTTHLTGDDNKKKLGKVKTKLNK-----TSANGNTATNFN 917
Db 575 TTNANTNTSNTATNTNTNSLTETV-----NLGEDALKWDKDKNGVFTAAHGTTETSKI 628
Qy 918 VNSSDEDALVNAKDIATENLTLAKEIHTTKGTADTALQTFVKKVDENNADANAIVG 977
Db 629 TNVKGDLTTGTDAVN-----GSQKLTNDAV-----ATNTNIATN 666
Qy 978 OKNANNQVNTLTGKENGNIKTDK-NG--TVTFQINTTSLGKAKSTLNDGGLSIKNP 1034
Db 667 TTNISLTETVTNLGEDAL--KWKDKNGVFTAAHGNNTASKI---TNLDCGTVT---AT 717
Qy 1035 GSEQI-----QVGADGV-----KEAKVN--NNGVVGAGIDGT-T 1065
Db 718 SSDALNGSQLDLSNIAITYFGGNASVNTDGVFTGPTYKIGETNYNVGDALAAINSF 777
Qy 1066 RITRDEIGFTGNSLDKSKPHLSKDGINAGGKITTNTQSIEAQNASHDAVGGIYDLK 1125
Db 778 TSLGDALLWDATACKFS-----AKHGTDASVITDVADEISDSSSDAVNGSQLHGV 831
Qy 1126 TELENKISSTAKTAQN-----SLHFSVADQGNNTFTVSNPYSSYDTSKTSVITF---AGE 1179
Db 832 SYVVDALGGGAENVADGTTITAPTYTIANADYN--VGDALNAIDTT-LDDALLWDADAGE 888
Qy 1180 NG-----ITTKVNGVVRVGIQDTKG---LTPKLTGVNNCKGIVIDSON 1222
Db 889 NGAFAAHGKDKTASVITNVANGAISAASSDAINGSQLYTTNKYIADALGGDAEV---N 944
Qy 1223 GQNTITGLSNTLIANY--TNDKGSVRTTEQGNIIKEDKTRAASIVDLSAGNLOGNEA 1280
Db 945 ADGITITAPTYTIANAEYNNVGDALDALLDNDLWDETANGGA-----GAYNASHDGKA 997
Qy 1281 VDFVS-----TYDTPNFADGNATTAKVYTDTSKTSKVYDV---NVDDTTIE---1325
Db 998 SIITNVANGSISEDSTDAVNGSQLNATNMI-----EQNTQIINLAGNTDATYIOENGA 1052
Qy 1326 ---VKDKKLGKVTTLTSTGTGANKFALSNOATDALV-----KASDIVAHLNLTSGDI 1376
Db 1053 GINYVTRTNDGLAFNDASAQGVGATAIGYNSVAKGDSVAIGQGSYSDVDTGIALGSSV 1112
Qy 1377 QT---AKGASQANNSAGYVDADGNKVIYDSDNKYIQAQNDGTVDKTRKAVAKDLVAQAQ 1433
Db 1113 SSRVIAKSRDTS-----ITENGVIYDITDGLLGLSIG-----1149
Qy 1434 TPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFVKLEKAASDNKTKNAAVTVGDL--1491
Db 1150 -----DDGKYRQIIN-----VADGSEAHDAVTVRQLQN 1177
Qy 1492 --NAVAQTPLTF--AGDGTAKKLGETLTIKGGOTDNTKLTDDNNIGVVAGT-----D 1540
Db 1178 AIGAVATPTTKYTFHANSTEDSLAVGTDSLAMGAKTIYN--GDKGIGIGYGAIVDANALN 1235
Qy 1541 GFTVKLAKDLNLSVNAGGTKIDDKGVSF-----VDSSGQAKANTPVLISANGLDLGGK 1594
Db 1236 GIAIGSNAQVTHVNSIAGISNGSTTTRGAQTYTAYNMDAPQNSVGEFSVGSAD---QOR 1291
Qy 1595 VISNVKGTQDITDAANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKPNSSSSNRT 1654
Db 1292 QITNVAAGSADTDAVNVGOL-----KVTDQAVSQNTQSIITN--1327
Qy 1655 VIKAGTVLGGKGNNDTEKLATGGIOGVYDKGDNANGDLSNVVVKTKQDGSKKALLATYNA 1714
Db 1328 -----LDNRVTNLDISRV--NIENGI-----GDIV-----T 1351

QY 1715 AGOTNLTNNPAEADIRNEQIRFEHVNDGNEPYYVQGRNGIDSSAGKSHSVAIGFOAK 1774
Db 1352 TGS TKFKTN-----TGVDASAQKDSVAI----- 1377
QY 1775 ADGEAAVAIGROTPAGNQSTAGDQQAIGTGNVYVAGKHSIGAIGDPSTVKAON 1834
Db 1378 -----GSGSIAAADN-----SVALTGTSV----- 1396
QY 1835 SYSVGNNOFTDQTDVFCVGNNTVTESNVALGSNSAISAGTHACTOAK-KSDGTAG 1893
Db 1397 -----ATEENTISVGSSTNORRITNVAAGKNA-----TDVNVQAQLKSSEAGG 1439
QY 1894 TTTTAGATGTVKGFAGTAVGASVGSAGAEIRIQNVAAAGEVSATSTDAVNGSOLYKATQ 1953
Db 1440 VRDTRADGSDID--YSNITILG-----GGNGTTRISNVAG-----VNNNDVVNVQAQLKQSVQ 1490
QY 1954 SIANATN-----ELDRHTHONENKANAGISSAMAMASMPQAIIPGRSMVTGGIATHNGQGA 2009
Db 1491 ETKQYTDQRMVEMDNKLSKTESKLSGSIASAMAMTGLPQAYTPGASMASIGGGTYNGESA 1550
QY 2010 VAVGLSKLSDNGOWPKINGSADTQGHVGAACVAGGFHF 2047
Db 1551 VALGVSMVSNRWRVYKLGSTNSQGEYSAAUGAGIQW 1588

RESULT 3
A82615
surface protein XF1981 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82615
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82615; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82615
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1190 <SIM>
A:Cross-references: GB:AF004017; GB:AF003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1981

Query Match 7.68; Score 778; DB 2; Length 1190;
Best Local Similarity 20.58; Pred. No. 8.2e-23;
Matches 429; Conservative 185; Mismatches 432; Indels 1044; Caps 72;

QY 64 SAYAQKDKTHIAICEQNPQRSG---TAKAD-----GDRA 96
Db 39 SAASSKGT-----QPRSNAMTAKRSKHLDRQLHIVLLAVLAASTGYTKVA 89
QY 97 IAIENANAAQCGQATATGSSNKT--VNGSSLDKTCGTATGQF-----SIAIGGDVKA 146
Db 90 AQVYVNSDSTENCVEILGSSQTSFIHSASNDKCKPDFQTQTEYSFLFYDRNLVLGG----- 145

QY 147 SGDASIAIGSDDLHLDDQHGNPKHPK-GTLINDLINGHVALKEIRSSKDNNDVKYRRTTAS 205
Db 146 -----SLYVNEGLGVDISGATYSMLGSIAT--MNGSAGIDSI-----ATGS 187
QY 206 CHASTAVGAMS---YAQGHFSNAFCFTRATAKSAYSIAVLGLAATAEGQSTIAIGSDATSSS 262
Db 188 GQGSTGDTGNTSATVAQGLRSIAIGTTAKRSQDRAISIGTGASTTGNFAIALGNALTSI 247
QY 263 LGAIALGACRQAQLAGQGSVVVTSQDNNSRPAYTPNTQALDPKFOATNNTKAGPLS 322
Db 248 ANGIALGASSSVYTRGGVALGQGSAAATASGIT--GYDPVTKSTSTLSMNRSTLGAVS 305
QY 323 IG-----SNSIKKKIINVAGVKNKTDVNVVAQLEAVVVKAKERRITFGQDDNSTDKVIGLD 378
Db 306 IGNITSSTQTRQLTCLAAGRSDTDAVNVAQLKLAE-----SVGGG 347
QY 379 NTLTKGGAETNALTDNNIGVVVKEADNSGLKVKLAKTLNLTENVTNTLNATTTVKVGSS 438
Db 348 WNLTAGSANSNVA-----LGESVDLKNSDGNLLIKT-----TDSNDVTENLATALKV--- 396
QY 439 SSTTABELLSLSTFTQPNTGSSQSTKTVYGVNGVFKFTNNAETAAITGTTTRTRDKIGFAR 498
Db 397 -----DSLTT-----TG--NTAMTTDGVTVGK-----RVTLDSTG--- 423
QY 499 DGDVDEKQAPYLDKKQKLVGSAITIDNIGIDAGNKKISNLAKGSSANDAVTIEOLKAAPK 558
Db 424 -----LVIAEGPSVVISSGAINAAGQIMNVGTADTDAVNFQLOAV-- 465
QY 559 TLNAGAGISVTPTEISVDKSGNVTAPTYNIGVKTTELNSDGTSDKFSYKSGOTNNSLVT 618
Db 466 -----SDTASKGWNLASGTSS-- 483
QY 619 AEHLASYLNEVNRDASALQSFTVKEEDDDANAITVAKDTTKNAGAVSILKIKGNGLT 678
Db 484 -----NVAPGASVDLKNTDGNL 500
QY 679 VATKKG--TVTFGLS---QDSGLTIGKSTLNNDGLTVKDTNEQIQVGANGIKFTNVNGS 733
Db 501 LITKAIGINDVTFENLATALEASLTTGNTAMTDTGVTV-----CS 540
QY 734 NPGTGIANTRTRDKIGFAGSDGAVDTNKPYLDQDKLOGVNVKVTNTGAINAGKAITGL 793
Db 541 N-----VT-----LGSTGLVITDGP-----SVTSSGISAGNQKIT-- 570
QY 794 SPTLPSIAQSSRNIELGNTIQDKDSNAASINDILNT--GNLKNNNPIDFVSTYDI 850
Db 571 -----NVAAGTA--DTDVAVNFSQLQAVSSSTASKGNL----- 600
QY 851 VDFANGNATTATVTHDTANKTSKVYVDVNVVDPTTIIHLTGTDNKKLGVKTTKLKTSANG 910
Db 601 --LASG-----AN-SSNVVPCESVD-----LKNSDGN-LLITKTIDSN----- 634
QY 911 NTATNFNVSSDEEDALVNAKDIAENLNTIAKEIHTTGTGADTADTALQFTVVKVVDENNADD 970
Db 635 -----DVTFNALATKLVDSLTTG-----NTAMT 657
QY 971 ANAITVQKNANNQVNTLTLLKGENGLNIKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSI 1030
Db 658 TDGVTVG-----SNTVLG-----STGL-----VTIDG--- 679
QY 1031 KNPTGSEQIOVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGTGTGNSLDKSKPHLSK 1090
Db 680 -----PSVTS 684
QY 1091 DGINAGKKTINTOSGEIAONSHDAVTGGKIYDLKTELENKISSTAKTQNSLHEFSVAD 1150
Db 685 SGISAGNQKITNVAAGTA---DTDVAVNFSQL-----QAVSSTASKGNL--- 726
QY 1151 EOCGNFTVSNPYSYDTSKTSVDITTEAGENGITTKVKNGVVRVGDIDQTKGLTTPKLTGVN 1210
Db 727 ASGANSNVAPGESVDLKNKD-----GN 749
QY 1211 NNGKGVIDSQNGONTITGLSNTLANVTNDKGSVRRTEQGNIIKDEDKTRAASIVDLVLSA 1270

Db 750 ----IVLSKESGSN-----||| 762
Qy 1271 GFNLGNGEAVDFVSTYDVFADGNATTAKVYDDTSKTSKVYVDVNDVDTTIVKDKK 1330
Db 763 -FNLSSS-----||| 773
Qy 1331 LGVKTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNLTSGDIQTAKASQANNSAG 1390
Db 774 LTVGDTVMVTNG-----||| 785
Qy 1391 YVDADGNKVIYDSTDNKYYQAKNDGTVDKTKREAVAKDLVAQAQTPDGTLAQMNKSVINK 1450
Db 786 -----||| 785
Qy 1451 EQVNDANKKOGINEDNAPVKLEKAASDNKTKNAAVTYVGDNLNAVAQTLPTFAGDTGTTAK 1510
Db 786 -----VTVG-----SGVTLG 795
Qy 1511 KLGETLTTKGGQTDNKLTDNNGVAVAGTGDFTVKLAKDLNLNSVNAAGTKIDDKGVSP 1570
Db 796 SMGLVI-----TDG-----804
Qy 1571 VDSSQAKANTPVLSSANGLDLGGKVISNVGKTKDQDAANVQOLNEVERNLLGLGNAGNDN 1630
Db 805 -----PSVTSSGINAGSKQITNVAAGTADTDVNLSQLN-----838
Qy 1631 ADGNOVNIADIKDPNSGSSSRNRTVIKAGTVLGGKGNNDTEKLPAT--GGIQGVVDKDCGNA 1688
Db 839 -----TAMAGSAGSKSVHYSTYDGGTQ-----GGNY 864
Qy 1689 NGDLSNWWKTKQDGSKKALLATYNAAQTNYLTNNPAEADIRINEQGIREFHVDNGQOE 1748
Db 865 NGD-----867
Qy 1749 PVQGRNGIDSSAGSKHSAVFQAKADGEAAVAGTQQA--GNQSIAGDNGAQTGDS 1807
Db 868 -----GATGTRIAVGVGTSLAEGATAVGSAAASGKSTAGIRNAVASADGS 916
Qy 1808 TAIGTG-----NVVAKHSG-----AIGDPSTVKADNSYSVGNQNFDTAT--1848
Db 917 VALDGAKGARGAESYTGKYSGLQNNVTGTVSGDASKEGTRTVSNVYDAKEATDAYNL 976
Qy 1849 -QTDVFG-----VGNNI-----TVTESNSVALGSNSA--ISAGTHA-----GTQAKK 1887
Db 977 RQLDRVAQDANRYVDNKTESLESEGQTFVKVNSL-----NNSATPIAAGVDATAIGVCATASG 1033
Qy 1888 SDGTA-GTTTATAGTGVKGFAGQTAVG-----AVSVGASGABRRIONVAAGEVSA 1937
Db 1034 ADSIANGKASASADNAV-----AIGNHSDVRANTVSVGSAGSERQVNTNVAAG----T 1083
Qy 1938 TSTDVANGSOLYKATQSIANATNELDRIHQENKANKAGISSAMAMAMPQAYIPGRSMV 1997
Db 1084 ADTDAVNSQL--NOGLITAKQYTDGVVGSILRRDQDGVAAAIATANLPQAYIPGRGMT 1140
Qy 1998 TGGIATHNGQCAVAVGLSKLSDNGOWPFKINGSADTQGHVGAAGVAGPHF 2047
Db 1141 SVGVSSYRCQSAIAVGVSSVSESGRWFKFGCSANTRSQVGTGAGVGYQW 1190

RESULT 4
B85547
probable rpx family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AB005174; NID:gl2513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

Query Match 6.2%; Score 641; DB 2; Length 5188;
Best Local Similarity 22.0%; Pred. No. 1.le-16;
Matches 519; Conservative 297; Mismatches 872; Indels 668; Gaps 113;

Qy 34 ATGQGVSVCTLSFARIAALAVLIVIGATLSSGSAYAKOKTKHIAIGEQNPRRSGTAKADG 93
Db 1436 ATAPSVTINTTADDDILNAAEAGSALTISGTAEAGQTVTVTLNGVNY---SGNVQADG 1492
Qy 94 DRAITAIGENANAQGOAIAIGSSNKTVNGSLDKITGTDATGOESIAIGDVKASGDASTA 153
Db 1493 SWSVSV-----PTGDLASLTASSVTVNASVSDKARNASATHNLTV--DLAAPVVTINT 1544
Qy 154 IGSDDLHLLDQGNPKPKGTLLINDLINGHAVLKEIRSKDNDVKYRRTTASCHASTAVG 213
Db 1545 VAGDDIIINATEHGQAO-----IISGSAT-----GATTGNTVSVTIG 1580
Qy 214 AMSY-----AOGHFSNAGFGRATAKSAYSLAVGLAATAEGQSTI-----ATGSDATSS 262
Db 1581 TTTYTTVLVDANGNWS--ICVPASV-----ISALAQGDVITITATVTDASNGSGTASH 1629
Qy 263 LGAIALGAGCTRAOLOSSI-----ALGQSVVTVQSDNNSRPAYTPNTQALDPKFOATNNTK 317
Db 1630 TVTVGALGAPVLAINITIAVDDIIINAAEAGADLAITGTSNOPAGQTIVTVLNGQNYTTTADA 1689
Qy 318 AGPLST-----GNSIKRKIIINVAGVKNKTDVANNVAQ 349
Db 1690 SGNWSVTVPASRVSAIGEATVTVTAATDADGNSGSASHNVQVNTALPGV----TINVVA 1745
Qy 350 LEAVVYKWK---ERRITFGQDDNDTDVIGLNDTITIKGAETNALTNDNNIGVVKREADS 406
Db 1746 TDDIIINAAEAGVEQTIISGQ---VTGAAGADTVTVTL-CGATYTTATVOANLSWSDVVPAS 1800
Qy 407 GLKVKLAKTLNLTETVNTTLNATTTVKVSSSSNTAEELSD-----448
Db 1801 AL-----GELGN-----GELTISASVTSVNGTNGTRETITIDANLPGLRVDTVAGDDVYN 1851
Qy 449 -----SITFQPNQTSQSTSKTVGVNGVKFTNN--AETTAAGTTRITRDKIGFARDG 500
Db 1852 IIEHQALVITGSSSGLAAGSNVTLTINCQTVAAVLADGTWSVG-----1896
Qy 501 DVDEKQAPYLDKKQLKVGVSVAITIDNGDAGNKKISNLAAGSSANDAVTIEOLKAAKPTL 560
Db 1897 -----VPAVDVSAMPAGSVTIAASGSTSAGNP-----VSVTHPVTVD--LSAVAVSI 1941
Qy 561 NA-----CAGISVTPTEISVDAKSGNVTAPTNYIGVKT-----593
Db 1942 NAITADDVINAKEGAALTLSGSTSGVEA--GQTVTVTF--GGKTYSATVAANGSNSTSV 1997
Qy 594 -----TELNSDGTSDKFSVKSGTNNSLVTAELHASYLNEVARTADSAEQSTVKEEDD 648
Db 1998 PAADMAALRDGDASAQASV--SNVNGNSATTTTHAYS-----VDASAPVTINTIAGD 2047
Qy 649 D-----ANAITVAKDTTKNAGAVSILKKGKNGL--TVATKKDGTVTFGLSODSGLTIG 700
Db 2048 DILNAAEAGAALTITGSSSTAEAGQTVTVTLNGTNTYGTGTGDSWSVSPSADLSLTAS 2107
Qy 701 KSTLN-----NDGTLVKDTNEQI-----QVCANGIKFTN-----VNGSNPG 736
Db 2108 XYTVNAAVSDRAGNPASVNNHNLTV--DTSVPVVTINTVAGDDVDVINATEHAQAQIISGATG 2166
Qy 737 TGIANTARITRDKIGIFA---GSDGAVDTNKPPLYDDKQLQVGNVKITNTGICINAGKRAITG- 792
Db 2167 AATGSTVTVTGTNTFTTVLDASGNWSVCPASVVSALANGVTVINASVTDAGGNGSAT 2226
Qy 793 ----LSPTLPSIADOSSRNIELGNTIQDKDKSNAASI---NDILNTGFLNKNNNPIDFV 845


```

Db 2227 HQVTNGLPTITFNA---ISGDNILNADEKQPLTISGSGTCLATGAQVTVTLNGHNS 2283
QY 846 STYDIVDFANGN-----ATTATVTHDTANKTSKVYVDVNVDDTT 884
Db 2284 ATTD-----ASGNWTLVTPVSDLAALGAQANYTVSASATSAAGNTASSQANLLVDSGLPDT 2339
QY 885 IHLTGDD---NKKLGKVT---KLNKTS-----NGNTAT-----NFNVSS 921
Db 2340 INTVAGDDIINAAEAGADQTSIGVTVRAAGDVTVTGLGNTYTATVQSNLSWSVSPTA 2399
QY 922 DEDALYNKADIAENLAKETHHTKGTAD-----TALQTFETVKVDENNADANA- 973
Db 2400 DLQALNG-DLTTTASVWANGNTGSGTRDTTIDANLPLGLRDTVAGDVIDVNSIEHQAL 2458
QY 974 -ITVGOKNAN-NOVNTLTLKG-ENGLNIKTDKNGTVFTGINTSGLUKAGKSTLN-DGGLS 1029
Db 2459 VITGSSGLNAGAVLTVTVNSVAYSATVQADGWSVGIIPAANYSAMPAGLTVEVDGSS 2518
QY 1030 IKNPTGSEQ---IQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTNGSLDKSKP 1086
Db 2519 ANNPVSVSHPTVLDLTAVALISINTVASDDVDVNAEKG-----2556
QY 1087 HLSKGINAGGKKTITNIOGIEIAQNSHDAVT-GGKIY-----DLKT-- 1126
Db 2557 NLTLGSGTSG-----IESGQTV-----TVTFGGKTYTASVAANGSWSVNPAADLATLP 2605
QY 1127 -----ELENKISSAKTQNSLHEFVNADEQGNFTVSNPYSSVD---TSKTSDVITFAG- 1178
Db 2606 EGAANVOASVSASGNSASATHAYSV-DASAPTLTI-NTIASDDILNAAEAGSPLTISGT 2663
QY 1179 ---ENGITTKVN-KGVVRVGDQTKG---LTPKLTVGNNGKGVIVD---SONGQNTIT 1228
Db 2664 STAETGQTVTVTLNGATYTTGTVDGWSVSVPTISALGALNASNYTVSATVNDKAGNPGS 2723
QY 1229 GLSNTLANVNDKGSVRTTQGNIIKDEKTRAASIVDVILSAGFNLOGGEAVDFVS--- 1285
Db 2724 ASHNLAVDTTAPVLTINTVAGDDIINDAEHAQAL-VISGTSAG-----GEXGDVVSVL 2776
QY 1286 ---TYDVNFADGNATKATVYDDTSKTSKVYVDVNVDDTTIEVKDKLGVKTTTITSTG 1342
Db 2777 NGKTYTTTLDASGNWS-----VGVPAADVTALG 2804
QY 1343 TGA-----NKFALSNQATGDALVKASDIVAHLNLTSGD---IQPAKASQANNSAGY 1391
Db 2805 SGAQITIASVSDRAGNSDDASRTVTVSLSAPVISINTIAGDDVINATEKGSDLA--LSGT 2862
QY 1392 VDAGCNKVIYDSTDNKYQAOKNDGTVDKTEVAKDKLVAAQAPDGTGLAOMNVKSVINKE 1451
Db 2863 SDQAPGTAITVTLNGQNYSAITDASGNWSVTP-----ASAVSALGEATYSVTSVNTN-- 2915
QY 1452 QVNDANKKOGINEDNAFYKLEKASDNKTKNAATVTVGDLNVAQAQTPLTIFAGDTGTTAKK 1511
Db 2916 -----AQHNSSTASHNVQNTALPGITINPVATDDIINASEAGSAQTI 2958
QY 1512 LGETLTITKGQDTNKLFDNNIGVAGTGTGTVKLAKDLTNLNSVNAAGGTIKDKGVSVFV 1571
Db 2959 SGQVTGAAAGSTVTVEL-----GGKTYTATVQADLSLWNVSVPAADWQ-----A 3001
QY 1572 DSSGOAKANTPVLSANGLDLGGKVISNVGKTKD--TDAANVOQLNEVRNLLGL---GNA 1626
Db 3002 LQNGELTVNASVTNA-----VGNFGSGTRDITDA-----SLPGLRDTVTA 3042
QY 1627 GNDNADGNQVNIAD- IKKDPNNGSSSNRTVIKAGTVLGGKNNDT-----EKLATGGIOGVV 1682
Db 3043 GDD-----VNIIEHAQAVITGSSG---FAAGTALTIVVNNQTYATVILANGSWSVGV 3094
QY 1683 KDXGNAGDLSNVWVKTKDKSKKALLATYNAAGQTNVLTN-----NPAEAI 1729
Db 3095 PAT-----DVSN-W-----PAGTLNITVSGANSAGTQTSITPLIPLVDLTAVALSNSITSD 3144
QY 1730 DRIN--EQGIRP-----FHVNDGNQEPVVOGRNGIDS--SASGKHSV---AIGFOAKAD 1776

```

```

Db 3145 DAINAAEKAALTLISGSGTSGVEAGQTVTVTFEGGKTYTTTVAANGSWSTVPAADLAALRD 3204
QY 1777 GEAAVAIGRQTAQGNOSTAIGDNAQATGDQSI-----AIGTGNVAGKHSGA--- 1823
Db 3205 GDASAQV-RVTNVN-----GNSATATHEYSVDSNAPVTINTIASDNIINASEAAGVT 3257
QY 1824 ICDPSTVKADNSYSV---GNNNQFTDAQTQDVFVGVN-NITVTSNSVALGSNSAISAGT 1879
Db 3258 VSGTSTAQTGQTLTVTLNGTNYQTT--VQTD-----GNSWLTSPASDLTALANNGYTLTAT 3311
QY 1880 ---HAGTQAKKSDG-TAGTIT-----TTAG-----ATGTVKG----- 1906
Db 3312 VSDLAGNLGSASKGVTVDTTAPVIFNTVAGDDVINVEHQAQIISLTATGAVAGDRLV 3371
QY 1907 --FAGQ-----TAVGAVSVGASGAERRIQNVAAGEV--SATSTDVANGS-----QLY 1949
Db 3372 VTIAQQQVYVSTDSAGNSKSVGPASV--ISGLADGTVTISATITDSAGNSSTQTQHNQVN 3429
QY 1950 KATQSIANATNELDHRHONENKANAGISSAMAMAPQAYIPGRSMVTGGIAT--HNQ 2007
Db 3430 TAAVSLSVSTISGDNLI---NAAEAG--SALTLSG-----TGTNFATGTVTVTVLLNGK 3477
QY 2008 GAVAVGLSKLSDNGOW 2023
Db 3478 GYSAT-----IQNSGSW 3489

RESULT 5
A:1477
190K surface antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
C:Accession: A41477
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
Infect. Immun. 58, 2760-2769, 1990
A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated,
A:Reference number: A41477; MUID:90354033
A:Accession: A41477
A:Molecule type: DNA
A:Residues: 1-2249 <AND>
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi
C:Keywords: surface antigen; tandem repeat
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Query Match 6.1%; Score 627; DB 2; Length 2249;
Best Local Similarity 22.2%; Pred. No. 1.3e-16;
Matches 477; Conservative 233; Mismatches 754; Indels 684; Gaps 108;

QY 97 IATGENA-----NAQGGQAIAGSSNKTVNGSSLDKIGTDATGQE---STAIGGDVKA 146
Db 44 IATNNNAAFSNVYNNNNWNEITAAGVAN-----GTPAGGQNNWAFYGGDYTV 92
QY 147 SGDASTAIAGSDDLHLDDQHGKPKPKG-----TLINDLI-NGHAVLKEIRSSKONDVKY 199
Db 93 TADAADRI-----IKAINVAGTT--PVGLNITQNTVGSIIITKGNLLPVTLNACKSLTLNG 146
QY 200 RRTTASGH-----ASTAVGAMSYAQHFSNAPGTRATAKAYSAYSLAVGLATAGQSTIAI 254
Db 147 NNAVAANHGFADPADNVTYGLGNIALGANAALIIQSAAPS-----KITLAGNIDGGGLITV 202
QY 255 GSDA-----TSSSLGAIAGAGTRAQLOGSI-----ALGGQSVVTSQDNNSRPAYT 300
Db 203 KTDAAINGTIGNTNALATVNVGAGT-ATLGGAVIKATTTKLTNNAASVLTITNANA----- 256
QY 301 PNTQALDPKFOATNNTKAGPLSTG-----SNSIKRKIIINVAGVKNKTDVAVNVAQL 350
Db 257 VLTGAID---NTTGGDNVGVNLNLGALSQVTDGIGNTNSLATISVAG-----TATL 305
QY 351 EAVVWKAKERRITFGQDDNSTDVKIGLDNLTITKGGAEETNALTDNNIGVYKADNSGLKV 410
Db 306 GGAVIKATTTKLT-----DAASAVK--FTNPVVVTGAIDNTGNANN--GIVTFICNSIVTC 357

```


Qy	411	KLAKTLNNLFEVNTTLNATVTTVKVGSS-----SSTTAELL--SDSLTTFQPWNGS	459
Db	358	NVGNT-----NALATVNVGAGLLOVQGVVKANTINLTNASAVTETNPVVVT	405
Qy	460	QSTSKTVYGVNG-VKFTNNAETTTAIGCTTRITROKIFARGDGDVEKQAPYLDKKOLKVG	518
Db	406	GAIDNTGNANNGLIVTFGTNSTVTGDIQNTAL-----ATVNVG	443
Qy	519	SVAITIDNG-IDAGNKKISNLAK-----GSSANDAVTIEQLKAAKPTLN	561
Db	444	AGTATLGGAVIKATTTKLTNAAASVLTLTANAVLTGAIDNTTGGDNGVNLNGLALSQVLT	503
Qy	562	AGAGISVTPETSEISDAKSNVTAPYYNICVKWTELNSDGTSDKFS-----VKGS-----GT	612
Db	504	GNIGTNTSLATISVGA--GTATLGGAVIKATTTKLTLDAAASAKFTNPVVVTGAIDNTGNA	561
Qy	613	NNSLVTAELHASLYLNVNRTADSALOSQTVKEEDDDANATTVAKDTTKNAGAVSILKLK	672
Db	562	NNGLIVTFGTNSTVTGDIQNT--NSLATISVGAGTATLGGAV-IKATTTKLTNAAASVLTLT	618
Qy	673	GKNGLTVATKKDGT-----VTFGLSQDSG-----LTIGKSTLNNDGLTVKD	713
Db	619	NANAVLTGAIDNTTGGDNGVNLNGLALSQVLTGDIQNTNSLATISVGAGTATLGGAVIKA	678
Qy	714	TNQIQVGGANGIKFTN---VNGSNPCTGANTARTIRDKIOFAGSDGAVDTNPKYLDQDK	770
Db	679	TTTKITNAVASAKFTNPVVVTGAIDSTGNANNGIYV---FTGNSTVTG-----	723
Qy	771	LOWGNKITTNGINAGCKAITGLSPTLPSIADSSRNTELGNTI---QDCKDKSNAASIND	827
Db	724	-DIGNTNALAT-VNVGAGTAT-----LGAVIKATTTKLTNAAASVLTLT	763
Qy	828	ILNTEGNLKNNNHPIDFVSYDIVDFANGNATTATVTHDTANKTSKVYVDNVDDTTIHL	887
Db	764	LTNANAVL---TGAIIDNTTGGDNGVNLNGLALSQVLTGDIQNTNSLATISVGAG-----	814
Qy	888	TGTDDNKKLGVTTKLNK-----TSANGN---TATNFNVNDSDEDALVNAK-----D	931
Db	815	TATLGGAVIKATTTKLTNAAASVLTLTANAVLTGAIDNTTGGDNGVNLNGLALSQVTDG	874
Qy	932	IABENLNTLAK-----EIHTEGTADTALQTFVKKVVDENNADANAI--	974
Db	875	IG-NTNSLATISVGAGTATLGGAVIKATTTKLTNAAASVLTLT-----NANAVLTGAIDN	927
Qy	975	-----TVGQKANNQVNTLTKEGNLIKDKDNGTVFTEGINTTGLKACKSTLNDGGLS	1029
Db	928	TTGGDNGVNLNGLALSQVLT--GDIG--NTNSLATISVG-----AGTATLT--CGAV	972
Qy	1030	IKNPSTGEQIOVGADGVKFAKYNNGNVVCGATDCTTRITRDEIGTGTN---GSLDKSKP	1086
Db	973	IKATT--TKLTDRAASAKR---TNPVVVTGAIDNTGNANNGIYVTFGTNSTVTGNVGTNA	1027
Qy	1087	HLSKDGINAGKKITNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHEF	1146
Db	1028	-LATVNVGAG---LLOVQGVVKANTINLTD-----NASAVTFNPVVVT	1068
Qy	1147	SVADEQGNFTVSNPYSSYDTSKTSVDITFAGENIT-----TKVKNGVVVRGIDQ	1197
Db	1069	GAIDNTGN-----ANNGIYVTFGTNSTVTGNVGTNALTATVNVG---AGLLQ	1111
Qy	1198	TKG-----LTPPKLTVG-----NNGKGIVIDSONGONTITG-LSNVT	1233
Db	1112	VQGVVKANTINLTDNASAVTFNPVVVTGAIDNTGNANNGIY---TTGTNSTVTGDIQNT	1169
Qy	1234	LANYTNDKGSVRTTEOGNIIKDEKTRAASIVD-----VLSAGENLQCNGEAVDF-----	1283
Db	1170	NALATVNVGAGITLQAGGSL-----AANNIDFGARSTLFEFNGPLDGGCKAPIYYFKGA	1222
Qy	1284	-----VSYDYDVNFADGNATTAKVYDODTSKTSKVVDVNVDD	1321
Db	1223	IANGNNAILNVNTKLLTASHLITGVAEINICAGNLFTIDASVGD-----VTILNAQN	1275

Qy	1322	TTIEVKKLGVKVTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNLTSGDIOT---	1378
Db	1276	INFRADSVLVSNL---TGVGVNNILL-----AADLVAP---GADEGTWVF	1316
Qy	1379	---AKGASOANNSAGV---DADGNK---VIYDS---TDNKYYQAKNDGTVDKTKEVA	1424
Db	1317	NGGVNGLNVGSNAGTARNIGDGGGNKFNTLLIYNNAVTTDDVNREGIOLVNLKNADET	1376
Qy	1425	K-----DKLVAQAOTPDGTL--AQMNVKSINKEQVNDANKKOGINE	1464
Db	1377	SSTAFNAGAIQINDATYTTIDANNGLNIPAGNIQFAHADAOQLVQLNSSGNDRTITILGANI	1436
Qy	1465	DNAFVKLEKAASDNKTKNAAVTVGDLNAV-AQTPLTFAGDTGTT---AKKLGSETLIK	1520
Db	1437	D-----PDNDDEGIVI---LNSVTAGKKLTIAG--GKTEGGAHKL-QTILFKG	1478
Qy	1521	GQDTNKLTDNIGVWAGTDGFTVKLAKDLNLNSVNAGGTIKDDKGVSVFYDSSGQAKA	1579
Db	1479	AGDCST-----AGTTFTNTNIVLIDIT-----QO---	1501
Qy	1580	NTPVLSANGLDLGGKVISNVGKGTCTDAANVOQLNEVRNLLGLGNAG---NDNAD---	1632
Db	1502	-----LEUG-----ATTANVFLFNDAVOLTQTGNIIGGLDFDNAGKNWV	1539
Qy	1633	--GNOVNIADIKDPNSGSSSNNRTVIKAGTVLGGKNNDTCKLA---TGGIQGVBDKDN	1687
Db	1540	TLNNNVNAGAVQ--NTGGTNGTLLI---VLGASNLNRVNGIAMLKVGAGNVTIAKGGK	1593
Qy	1688	AN-GDLSNVVVKQDKGSKKALLATVYNAAGQTNLYLTPNPAEADIRNEOGIRFHHVDGN	1746
Db	1594	VKIGETQGTGTWT-----LTLPAHFNLTG-----SINKTGGQALKLNFNNGS	1636
Qy	1747	QEPVVGGRNGIDSSAGKSHSVAIGFOAKADGEAANAIVAIGRTQAGNOSIAIGNDAQAT---	1803
Db	1637	VSGV-----GTAAANSVGDIITTAGATSFASSVNAKGTATL	1671
Qy	1804	-GDQSTIA---IGTGNVAVACKHSGAIGDPSTVKADNSYSVGNNOQFDTATQDVFVGNNI	1859
Db	1672	GTTTFEANTFTNTGAVTLAK-----GSITSFAKNVTATSE-VANSA	1711
Qy	1860	TVTESNSVALGNSAISAGTHAGTQAKKSDGTAGTTTTAGA-----TGTVKGFAGQTAVG	1914
Db	1712	TINFNSLAFNSN--ITGG-----GTTTLILGANQVTVTGT--GSFTDITLNL	1754
Qy	1915	AVSVCAG--SGAERRQNVAAGEVVSATSTDAVNGSQLYKATQSTANATN	1960
Db	1755	TFDGAAGKSGGNILKSGSLDLSGVSTLAL-----VVTATN	1791

RESULT 6

B71704

cell surface antigen (sca3) RP451 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

```
C;species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence revision 21-Nov-1998 #text change 03-Nov-2000
```

C;Date: 21-Nov-1998
C:Accession: B71704

C;Accession: B71704

R; Andersson, S.G.E.; Zomlefer, S.B. 1986. The effects of...

Nature 396, 133-140, 1998

A;Title: The genome sequence of *Rickettsia*

A; Reference number:

A:Reference Number: A/1630; M01D:99039499
A:Accession: B71704

A;Accession: B/I/U4
A;Status: preliminary: n

A: Status: preliminary; no
A: Molecule type: DNA

A; Molecule type: DNA

A;Residues: 1-2340 <AND>

A; Cross-references:

A; Experimental sour

Query Match

Query Match	3.6%	Score 382;	DB 2;
Best Local Similarity	21.2%	Pred. No. 7.2e-15;	

Best Local Similarity 21.28; Pred: No. 7.2e-15;
Matches 499; Conservative 313; Mismatches 879;
Indels 659; Cans 110.

Db 2155 VYGRNKKIKYKRIITITDQIAIGKFINFYSEYELGYNLYIS---HRTTITPMFGNRY 2211
Qy 1894 QAKSDGTAGTTTATAGTGVKGRAG--QTAVGAVSGVSGAERRIQNVAAGEVSAITSD 1941
Db 2212 ATFRNNGYKENNTTFQNLISIKKYYDKFTILGLNSV-----THYLSQD 2255
Qy 1942 AVNGSQL-----YKATOSIANATNELDHIHONENKANAGISSAMAMASMPQAYIPGRSM 1996
Db 2256 IIIPELHWFNYOCKNKLPNIDARLD-----GIDEPLTITRFPKPAKITYN-- 2301
Qy 1997 VTGGIATHN 2005
Db 2302 LGGGISTKN 2310
RESULT 7
T31105
hypothetical protein 2 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31105
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326
A:Accession: T31105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4919 <NAR>
A:Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
C:Genetics:
A:Gene: lspA2
Query Match 5.5%; Score 570.5; DB 2; Length 4919;
Best Local Similarity 19.2%; Pred. No. 5.3e-14;
Matches 484; Conservative 353; Mismatches 914; Indels 771; Gaps 109;
Qy 2 NHIVKVPNKATGTFMAVAEYAKSHSTGGGSCATQGVGS-----VCT----- 43
Db 3 NKRYKLIFS KVKVKNCLVPVVAENIKSASGSSGSSSKIAEDQEEPEPDSIACSLPSSSIH 62
Qy 44 -----LSFARIAALAVLIGATLGSAYAKKDKTHIAIGEONQPRRS- 86
Db 63 LGLNHSPLKVPKGNUSVLLSLMPAQVWADSSNAIVDHSAGAKQAVDERD-PKNGK 121
Qy 87 -----GTAKADGDRAI-----AIGENANAQG-GQAIAIGSSNKTGVNGSSLD 126
Db 122 EKVVVINIAKPD-EQGISDNHFSKFNIPNSAVFNNSIKEGNSQLVGLLGENKNLGSQAQ 180
Qy 127 KIGDTATQESIAIGGVKASGDASIAIGSDDLHLDDHGNPKHPKGPLINDL--INGHA 184
Db 181 TIFNOVTGDOESKISGGLEVFGEKA-----DLFTI-----NPNGVTLNGVKTINTDR 227
Qy 185 VLKEIRSKDNDVK-----YRTTASGHASTAVGAMSYAQGHFSNAFGTRATAKSAYSLAV 240
Db 228 FVASTSEVVEPHIKOLANVQRKVIIGDKGVATNGLS-----HFD----- 266
Qy 241 GLAATAEQSTIATGSDTSSSLGAIAGCAGTQAQLOGSIALGOGSVVVTQSDNNSRPAYT 300
Db 267 VVARNIEQGVKSGEGDSKPAKLANVTAAGNLIY-----DVNTRDV-N 309
Qy 301 PNTQALDPKFOATNNTRKAPISIGSNS-----IKRKIINVAGAVN-----KTDVAVN 346
Db 310 RNT---NPKKPTDNTNRKDNIATISGESAGSMYGRNIKEPIVTDKGAGVNHQGVIFAEDDIN 366
Qy 347 V-----AQLFAVYKWKAKERITFQGDNDNSTDVKIGLNTLTIKGGAEINALT 393
Db 367 ILTDDGNSRLNKKVYADYVRVY--GKDIELANNQIHA--DOQLILNATGHVKNLGDGSSVIS 423
Qy 394 DNNIGV-----VKEADNSGLKVKLAKTLNLTENVNTTTLNATTVTKVVGSSSSTA 443
Db 424 NNLGISALNTLENATVSNANLSFRVTNDTKLNLKSVKARSARAAD-----LQSGNLNLDKA 479

Qy 444 ELLSDSLFTQPTNGSQSTKTVYGVNGVKF-----TNNAETAAICTTTRITRDKIGF 496
Db 480 SVLAHKLTLNISNDVSLN-NOSKLSANNLKIKKVRDLNLSNLSA-----NNLT 529
Qy 497 ARDGDVDEKQAPYLDKKQKLVGSVAITDNGID-----AGNKKISNLAKGSSANDAVT 549
Db 530 NTSNNITLK-----NKSFTAGMTLNTVNNVTLNNDSELANNTLNTVTRKNTLND-- 582
Qy 550 IEOLKAARPTLNAGAGISVTPTEISVDAKSGNVTAAPTNIIG-----VKTTELNSDGTSD 603
Db 583 -SKLSANKLDLN-----VTDNVTLSKS-----TLSAGELTFKKVKNVTLNDSE-- 626
Qy 604 KFSVKGSGTNNLSLVTAEHLASYLNEVN---RTADSALQSFTVKEEDDDANAITVAKDTT 660
Db 627 -----LAANNLSLNASHNVTLLNNKSKLSAQKADIKAVNLTLNDTTELTAKNLDINSTTI 680
Qy 661 KNAGAVSIL-----KLKKG-----NGLTVATKKGDTVTFGLS 692
Db 681 TNGGTIAGIFANITTEKLNKKEKALILAEQNLNFTVNGSHYENKGDIVSKDKATVTFSKN 740
Qy 693 QDSGLTICKSTLNDGLTVKDTNEQIOVGANGIKFTNVNGSNPGTGCIANTARITRDKIGF 752
Db 741 SD-----FTSNGSKLVNAQNLKVNVTISQ-----GDDITLIGNVTNLNASGT 785
Qy 753 AGSDGAVDTNKPYLDQDKLQGVNKK-ITNTGINAGGKAITGLSPTLPSIADQSSRIELG 811
Db 786 FTNSGNLTIVK-----TLDVGDIIQNFNTKG-----NLTVG 815
Qy 812 NTIQDKKS-----NAASINDILNTGF-----NLKNNNPIDF 844
Db 816 EDLHKSKTKITNDGKLISIKNLNISSEADFINNLTGLIEALKIATKGNFTNKEKAILA 875
Qy 845 VSTYDIVDFANGNATPATVTHDTA---NRTSKVVYDVNVDDTTH-----LGTDDN 893
Db 876 SNLLDLSVAEGKKTFFNGTIESGKNLNTITNGAF-LNVDNATIRSGVLNITSTGNVSN 934
Qy 894 KKLGVKTKLKNKTSANGNTATNFVNSSDEDAVNAKDAENLNLTAKEIHFTKGTADTA 953
Db 935 NGTLISNERLNTS-----AANF-TNESNGTVMSNGL-----LNIIAQGNITNKNLIAS 983
Qy 954 LOTFTVKVDEN--NNADDANAITVGQK-----NANNQVNTLTKGE-----NGLNIKT 1001
Db 984 RQQLMLTAVADNITNDSNISKNIAVLHLSLGNISLSKDVYNL-----GEIYANNISVKAH 1040
Qy 1002 --KNGTVFGINTTSGLKAGKSTLNDGGLSIKNPTGSEQIOVGADGVKFAKVNNGVYGA 1059
Db 1041 QLKNDVKLMDITTK-----TREGQASYKLYQASNGHFGNDGSSGYSBGDLNKGK 1092
Qy 1060 GIDGTTRITRDEIG--FTGTNGSLDKSKPHLSKDGINAGK-----KIT--- 1101
Db 1093 FADLNKLTIVORIGKIYAGROLTFNKS-----NAGGSEIINRGITNVNKNLSYDS 1143
Qy 1102 -----NIOGELIAQNSHDVATGGKIYDLKTELENKISTAKTAQNSLHFSVADEQGN 1155
Db 1144 DVSFENNQOSKV-----DLYT--KIFEAKSDIE-----LTFKT--NGTHPVYL-----N 1184
Qy 1156 FTVSNPYSSYDTSKTSVDITFEAGE---NGITTKVKNKVVRVGVIGDQTKGLTTP---KLTVG 1209
Db 1185 PKSNNEKKYRSENTEKPKSIGDLINEALSADSAPEAEYAYSGSSSNVINPVSYLAALG 1244
Qy 1210 NNGKGIVIDSONGQNTITGLSNTLAN-----VTNDKGSVRTTE 1248
Db 1245 NAN-----NSSPHYLNTALKHILGNQWQDLKKOENIKVLKQWEDPKDKGASKMLD 1298
Qy 1249 OGNIIKDEKTPAASIVDVLSAGFNLOGNEAVDFVSFYDTVNFADG---NATTAKVYD 1305
Db 1299 ---LTPNTDKKAKIFAGIIRNGNDTISDVESEDFKKYS---KFQNGEAKWAKDGTDSYD 1353
Qy 1306 DTSKTSKVYDVNVDDTTIEVKDKKIKGVKTTTLTSTGTGANKFALSNOATGDAVKASDI 1365
Db 1354 STKASEKYYKVENVDHKE-NIDEHKLNGKHEITVPGVSFEN--LNNKNMDHQDPKLGEI 1410

QY 1366 VAHL--NTLSGGDIOTAKASQ-----ANNSAGYVDADGNKVIYDSTDN-----KYYQAK 1412
DB 1411 DKSIISELLAQPVYTEKSAARSDPRVNDKEDALDNLNRYTRLISYINQNNYLGAKEYFNQ 1470
QY 1413 NUGTVDKTEVAKDKLVAQAQPDGCTLAQMNVKSYLNKEQVNDANKKOCINEDNAFVAGL 1472
DB 1471 LDPEDDKLKGIR-----IGDNFEHQILTRLIEKVADNHLHLKGLH-DIALVKL 1521
QY 1473 EKAASDNKTKNAAVTVGDNLNAVAQTLPLTFAGDTGTTAKKLGETLT----- 1517
DB 1522 IDSAS-----IQAKDLNL-----KVGEALTKEQDKDLKDLVWV 1556
QY 1518 -----IKGGOTDKNKLTDNNIGVYAG-TDGF 1542
DB 1557 KTEVNGQEVLPVYLAQOTIEVEKORGVGCTQIRAGIIDVKKDDVRNTGTIAGVAGL 1616
QY 1543 TVK-----LAKDLTNLNSVNAAGTKIDDKGVFVDSG----- 1575
DB 1617 EAKNKLNTGDIILSORLSK-----VGKGLSTGVTVYVDETGATKVRKARIKSEGHILE 1672
QY 1576 -----QAKNTPVLSANGLDLG-----GKVISNVGKTKDTDA 1608
DB 1673 TDKDKNVDLTASELKGNTGQIKAKDLNLDIYETSYKYKELFGKNGGEIGDRVTQTSQ 1732
QY 1609 ANVQQLNEVENLLGLGNAGNDADGNQV-----NIADIKKDPNSGSSNR-----TVIKAG 1659
DB 1733 AKSVGTDAFPHLHLSLEGVDVNOTGSLKANRTTGVVKGDFNTKAGKDLFHRQIDVTSG 1792
QY 1660 TV-----LGGKGNNDTEKLTGGIOGVVDK--DGNANGDLSNVWVKTK----- 1701
DB 1793 TVYSASAGGQSGAGISLTDQGVETVNTKTATAGANADVTFNFKRTRETETSLTHRNSEF 1852
QY 1702 -----DGSKKALL-----ATYNAAGQ 1717
DB 1853 NALSGELYVMCKADIGVDINRDVEVIKTPPEETAABQKAEBAKAEVENEASETAKE 1912
QY 1718 TNYLTNPAEAIIDR-----INEQIRFFHVNDGNQEPVVGQNRGIDSSA 1761
DB 1913 TEEAENDVAEKDKTKPKFKKLTDEETAAAFETKGEDEFFAAYKAREEE--DRKKGETLSA 1970
QY 1762 SGKHS-----VAIGFOAKADGEAAVAGIRQ-----TQAG--NOSTAIG 1797
DB 1971 EQIESTKARDEKETYVELKVGVAEAEHSAADAISNKARQIIDFQNGKQGGVVALQ 2030
QY 1798 DNAQ-----ATGDQSIATGTVNVAGKHSAGIDPSTVK-----ADNSYSVGNNOQFTDAT 1848
DB 2031 EASDVNLATGD-----LAGASAKLKFELSTIEKKSARGASDCRSILGGRLLAAR 2080
QY 1849 QTDVFGVGNNTVTSNSVALGSNSAISAGTHAQTKKSDGTAGTTTAGATGVKGA 1908
DB 2081 GGDII--TLNNVETTENSLSLKARDNVNVNSGV-TEQKDESNOSLKVTAGASSGCGVMA 2137
QY 1909 GOTAVGVSAGSAGERRIQNVAAAGEVSATSDAVNGSQL-YKATQ--SIANATNELDHR 1965
DB 2138 GCCSAG-VSAGVSGS-----YNESNTESHTNSLLRGLSRVEAGKDFNLISSNVVDH- 2191
QY 1966 THQENKANAGISSAMAMAPQAYIPGRSMVTGGIATHNG---QGAVAVGLSKLSDNGQ 2022
DB 2192 LHL-DVKGDTNVVKQDSYSRKRERGVNYSVAGVGVSTAGGARPNGSVGLGVSAENENSK 2250
QY 2023 WV 2024
DB 2251 IV 2252

RESULT 8
E85822
probable invasin 23135 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <STO>
A:Cross-references: GB:AE005174; NID:q12516151; PIDN:AAG57041.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3135

Query Match 5.1%, Score 521; DB 2; Length 2660;
Best Local Similarity 20.8%, Pred. No. 1.9e-12;
Matches 484; Conservative 277; Mismatches 880; Indels 682; Gaps 103;
QY 23 AKSHSTGGSCATGQVSGVCTLSFARIAALAVLIVIGATLSGSAYAKKDKTHIAIGEONQ 82
DB 546 ADSHST-----ATLTFIAHDAAGNPVIGLVLS-----TRH----- 575
QY 83 PRRSCTAKADGDRAIAIG---ENANAQGGQAIAGSSNKTVTNGSSLDKI-GTDAFGQBSI 138
DB 576 -----EGVQDITLSDWKDNGDSYQILTTGAMSGTL--TLMPQLNGVDAKAPAV 624
QY 139 AIGGDVKAS-GDASTAIGSDDLHLDDQHGNPKHPKGTLLINDLINGHAVLKEIRSKDNV 197
DB 625 VNIISVSSSRTHSSIKDKDRY---LSGNPIE-----VTVELRDENDKPV 666
QY 198 KYRRITASHASTAVGAMSYAOGHFSN---AFGTRATAKASAYSLAVGLAATA-----E 247
DB 667 KEOQ-----OLMTAVSIDNVKPGVTDMKETADGVYKATYATYTKGSLTAKLLMQNNE 722
QY 248 GOSTAIGSDATSSSILGATGATRAOQLOGSIALGQGSVVVTQSDNNSRPA--YTPNTQA 305
DB 723 DLHTAGFIIDANPQSAKIATLSASNNGVLANENAANTVS-VNVADEGSNPINDHTVTFVA 781
QY 306 LDPKFOANTWTKAGPLSIG-----SNSIKRKIINVCAGVYKNTDA 344
DB 782 LSGSATSFNNQNTAKTPDVNGLATFDLKSQEDNVVEVTLENGVQKTLIVSFGDSSTAQ 841
QY 345 VNV-----AOLEAVVWKAK-----ERRITF-----QGDDNS----- 370
DB 842 VDLQSKNEVWADGNDSDATMTATVRDAKGNLLNDVKVTENVNSAAAKLSQTEVNSIDGIA 901
QY 371 ----TDVKIGLDNLT--IKGAEETNALDNNIGVYKEADNSCLKVKLAKTLNNLEVTNT 424
DB 902 TAPLTSKNG-DYVTVASVSSGSAQ-----QVIFIGDQSTAAULTSVSPSGDIIVTWT 954
QY 425 TTLNATTTVKVSSSTTAELSDSLTQPNQTSQ--STSKTVYGVNGVYKFTNNNAETTA 482
DB 955 APLHMTATLQDKNGN---PLDKKEITFSPNDVNASRFSISNKGKMTDSNGTAIASLTG 1010
QY 483 AIGTTTRITRDKIGFARDGVDKQAPYLDKKQLKY---GSVAITDIONGIDAGNKKISNLA 539
DB 1011 TLAGTHMITARLANSVSDT-OPMTFVADKRAVVVVLQTSKAEIINGVD--ETTITATV 1067
QY 540 KGSSANDAVTIEOLKAAPTLNAGAGISVTPTEISVDAKSGN-----VTAPTYNIGVKIT 593
DB 1068 KDPFDN---VVKNLVSVFRT-----SPADTQLSLNARNTNENGIAEVTLKGTVLGVHT 1117
QY 594 TE-LNSDGTSDKFSVK--GSGTNNSLVTAHLASYLYNEVNTADASLOSFTVKEDEDDAN 651
DB 1118 AEAILLNGNRDTKIVNIAPDASNAQVTLNIPAOQV--VTNNSDSVQLTATVKDPSNHPVA 1175
QY 652 AITV-----AKOTTKNAGAVSILKKGKNGLTVATKKDGT----- 686
DB 1176 GITVNTFTMPOVAAVFLENNG-IAITQANGEAHVILKKGKAGHTVTTATLGNNSDAQ 1234
QY 687 -VTFGLSQDSGLAT---GKSTLNNNDGL-----TVKDTNEQIQOVGANGIKFTNVN-GSN 734
DB 1235 PVTFVADKDSAVVVLQTSKAEIINGVDDETTLTATVKDPFD-----NAVKDLQVTESTN 1288

Qy	735	PG-TGIANTARTRD-----KIGFAGSDGAVDTNPKYLDQDKLVGVNKKIWTNIGINAGGKA	789
Db	1289	PADTQSSQSNTSGSVAEVTFKGVLGVHTABEATLPNGN-----NDTKIIVNADIAPDSNAQ	1345
Qy	790	ITGLSPTLPSIADOSSRNIELGNTTODKDKSNAASINDILNTGFNLKNNNNPTDFVSTYD	849
Db	1346	VT-LNIPAOQVNTNNSDSVOLUTATVKDPSNHPVAGIT-----VNFMPQDVAANFT	1395
Qy	850	IYDFANGNATTATVTHDTANKTSKVYDVVRVDDTTHLTGTDNKKLGVKTKLNTKTSAN	909
Db	1396	L-----ENNGIAITQANGEAHVTLK-----GKKAGTHTV-----	1424
Qy	910	GNTATNPNVNSDEDALVNAKDIAENLANTLAKEIHTTKGTADTALQTFVTKKVDENNAD	969
Db	1425	--TATLSNNNTSDSPVTFVAD-----KTSALVVLQISKNEITGNGV	1464
Qy	970	DANAITVVGOKNA-NQOVNTLTLLKGENGLNIKTDKNGTVTFGINTTSG--LKAGKSTLNDG	1026
Db	1465	DSATLTLATVKDQDFNEVNL-----PVTFG-TASSGLTUTPGESNTNES	1507
Qy	1027	GLSIKNPTGSEQIOVGADGKFAKVNNNGVVGAGIDGTRTRIDREIGFTGTNGSLDKSKP	1086
Db	1508	GIA-----QATLAGVAF-----GBOVTAS-----LANNASDNKTV	1539
Qy	1087	HLSKDGINAGGKKITNIOGETA-----QNSHDAVTGKGIYD-----LKTELENKISSTAKT	1138
Db	1540	HFIGDAAAAKIETLFPVPSIIAGTPQNSGSVITATVVDNNGFPVKGVTVN-FTSNAAT	1598
Qy	1139	AQNSLHEFSVADEQGN-NFTVSNPSSYDTSKTSDVITFACENGITP-----KVNKGVVR	1192
Db	1599	AEMTNGQAVTNEQGXATVYTNTRSSIESGARPDVBEASLENGSSLTSTSYNADAOST	1658
Qy	1193	VGIDQTKGL-----TTPKLTGVNNNGKVIDSONGQNTITGLSNTLANVTNDKGS	1243
Db	1659	AHLTLLOALEDTVAGSDGTTNLIYKDNNGV-----POEVTLSVSPEGVTPSNNNA	1712
Qy	1244	VRTTEGNI IKDEDKTRAASIVDVL SAGFNLOGNEAVDFYSTYDTVNFADGNATTAKVT	1303
Db	1713	IYTT-----NHDGN-----FVASFTAKKAGVYQVTTATLEN	1742
Qy	1304	YDTSKTSKVYDVNVDDTTIEVKDKLCV-----KTTTLTSTGTGANKFALSNOATGDA	1358
Db	1743	XDSMOQT--VTYVPVYANAEISLAASKOPVJANNNDLTTLTATVADTEGNAIAN-----	1794
Qy	1359	LVRASDIVAHLNTLSDIGOTAKGASQANNAGYVDADGNKVIYDSTDNKYVQAKNDGTFVD	1418
Db	1795	-----SEVTF--TLPEDVR-----ANFTLGDGCKGVVDT-----EGRAKVTLK	1830
Qy	1419	KTKEVA-----KDKLVAQAO-----TPDGTLLAOMNVKSVINKEQVND	1455
Db	1831	GTRAGAHTVVTASMAGKSEBQLVNFVFIADTLTAQVNLNVTEDNFIAN-NVGMTRLQATVTD	1889
Qy	1456	ANKKQGINEDNAFVKGLEKASDNKTKNAAVTVDLNAVAOTPLTFAGDTGTTAKKLGUET	1515
Db	1890	GNGNPLANEAVTFTLPADVYASFTLLCQSGS-AITDINGKAEVTL-----SGTKSGTYPVT	1943
Qy	1516	LTTKG-GQDFTNK-----LTDNNGVYVAGTGDGFTVKLAKDLTNLNSVNAGG	1560
Db	1944	VSNNYGVSDTKQVTLIADAGTAKLASLTSYFVSVSTEGAT--MTASVTDANGNPVEG	2001
Qy	1561	TKIDDRKGVSEFVSSGQAKAN-----TPVLSANGLDLGGKVISNNGKGTGD	1605
Db	2002	IKVNFRTGTSVTLSTSVETDDRGFAEILVTSTEVGLKTVSASLADKPTEVLSRLLNKAD	2061
Qy	1606	TDAANYQOLNEVRNLLGLG-----NAGDNDNAGNOV-----	1636
Db	2062	INSATITSLIPEGGYVMVAQDVAVKAHVNDQFNPILNESVTFSAEPPPEHMTISQNTVST	2121
Qy	1637	---NIADIKKDPNMSGSSNRTYIKAGTVLGGKGNND---TEKLTATGGIQ--VGVDKDGN	1687
Db	2122	DTGIIAEVMTWTPRNGS---YMWKASLANGSSEYKDLVVIDQKLTLSASSPLGICNGSPTG	2178
Qy	1688	ANGDLNVMVYTKQDKGSKKALLATYNAAGQTNLYLTNPNPAEAIIDRINBOGTRF-----	1739

RESULTS

RESULTS
576109

576109 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C; Species: *Synechocystis* sp.

A; Variety: PCC 6803

```
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
```

C;Accession: S76109

R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O, K.; Okumura, S.; Shimo, S.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A; 5

5

A; Reference number:

A;Accession: S76109

A;Status: preliminary

A; Molecule type: DNA

A;Residues: 1-3029 <KAN>

A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10087.1; PID:

Query Match 5.0%: score 518: DB 2: Length 3029:

Query Match	Score	Pred. No.	DB
Best Local Similarity	21.88%	3e-12	DB
Best Local Similarity	21.88%	3e-12	DB

DESC LOCAL SIMILARITY 21.0%, FREQ. NO: 35 12,
Matches 521: Conservative 256: Mismatches 901: Indels 710: Gaps 119:

56 VIGATL-----SGSAYA-----OKKDTKHTAIGE-----QNOPR 84

QY	DO	VI	OR	TH	SA	Q	RA	D	R	H	A	I	A	L	O	Q	N	Q	F	K	04
												:	:	:	:						

DD / VSOA13NVS0503SV111V1V1GTANNNG1VNEDFA3GQNIINDIANALSN1IF1IDEQ11 088

0V 85 RSGTAKADG---DRAIAIGENAN-AOGGOAIAI-----GSSN--KTVNGSSLDKIGTDAT 133

Db 687 LDNTAPAASTTLDANITADDIINIAESGOAIPITGTVGGEEFNVGDTVTLT.VVNDKTFMGAV 746

007 EDNIRFRRCTI DZRNLT R8DDTINIALESQQALFTLIGTVSGEFNVODI VILIVNDRTIFGAV / 48

QY 134 GQESI ---AIGGDVAKASGDASTAIGSDDLHLELDQHGNFKRHPKGTLEINDLINGHAYLKEI 189

Db 747 GAGGLESINVPGSDLIVDADLTIAA---SIATTDAAGN-----I. 782

0y 190 BSSKDNV KYRRITIA-----SCHASTAVGAMS----- 216

```

QY 130 KSRDNDVAKRRIA-----SGHASTAVGAMS-----Z10
      | : | | | | | | | | | | | | | | | | | | | |

```


QY 259 TSSSLGAIAG-----AGTAAQLQG-SIALGQGVVTTQS-DNNSRP-----AYTPT 303
Db 903 PIVTINIAVDDIINAVEAGSPVSGITTTGVEDGVVTTIDGNTYATVTGNWTFNI 962
QY 304 QALD-PKQALDNTNKTAGPLISGNSIKRKIINGAGVKNKTDVAVVQAEVAVVKKAKERRI 362
Db 963 PVADIANFEATEEVVATVSDLAGNPATPRNI-----TVDTVAPA-----V 1004
QY 363 TFQGDNDSDVK-----IGLDNTLTITKGAEATNALTDNNIGVVVKEADNSGLKVLAKTLNN 418
Db 1005 TIDSISDDTGAQANDFITNDDTLVNGTAEADST-----VVVSLDG----- 1045
QY 419 LFEVNTTTLNATTTVKVSSST-----TABLLSD-----SLATPQNTQSQSTKTV----- 466
Db 1046 -IEIGTVTAN-----GAGETLDYGTLLADGDYELSVTATNPTGNSATATQTVVDT 1097
QY 467 ----YGVGVKFTNNAETTAAGITRITRDKIGFARDGDVDEKQAPYLDKKQLKVGSVAI 522
Db 1098 TAPTVTINIAVDDIINAVEAGSPVAVSGTTTG-VEGQV-----VTV 1139
QY 523 TID-----NGIDAGNKKISLAKGSSANDAVTIEQLK-----AAKPTLNAGAGISVTPT-- 571
Db 1140 TIDGNTYATVTGNWTFNIPVADIANFEATEEVVATVSDLAGNPATPRNITVDTTAP 1199
QY 572 ----EISVD-----AKSGNVT-----APTNYGVKTTTELNSDGTSDKFSVK 608
Db 1200 TVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTTIDGNTYATVTGNWTFNIP 1259
QY 609 GSTNNSLVTAHLSYLNENRTADSALQSTFKVEEDDDANAIV-----AKD 658
Db 1260 VADIANFEATEEVVATVSDLAGNPATPRNITV-----DTVPAVTTIDSISDDTGAQAND 1315
QY 659 TTKNAGAVSLKLGKNGLTATPKDGTTFGLSQDSGLTIGKSTLNNDGLVAKDTNEQI 718
Db 1316 FITNDDTLF-----NGTAEA--DSTVVVSL--DGIEIGTVTANGAGETWLDYGTGL 1363
QY 719 QVANGIKFTNVNGSPNGICIANFARITRDKIGFAGSDGAVDTNKPYLDQDKLOVCNVKI 778
Db 1364 L--ADGDYELSVTATNP-TG--NSATATQTV-----VDTAPTVTINIAVDDI-- 1408
QY 779 TNGTINA--GGKAITGLSPLPSIAQSSRNIEL-GNTIQDKDKSNAASIN-----DILNT 831
Db 1409 ----INAVEAGSPV-AVSGTTTGVEDGVVTTIDGNTYATVTGNWTFNIPVADIAN- 1462
QY 832 GFNLKNNNPIDFVSTDIIV-----DFRANGNATTAT--VTHDTANKTSKVYDVVND-- 882
Db 1463 -----FEATEEVVATVSDLAGNPATPRNITVDTTAPT-VTINIAVDDIIN 1509
QY 883 ----TTIHLTGTDNKKKLGKVTTKLNKTSANGNTAT-----NPNVNSD----- 922
Db 1510 AVEAGSPVAVSGT-----TTGVEDGVVTTIDGNTYATVTGNWTFNIPVADIANFEAT 1565
QY 923 EDALVNAKDIAENLTLAKEIHTTKGTADTALQTFVTKVVDENNADANAITVGOKNAN 982
Db 1566 EEVVATVSDLAGNPATPA-----TRNITVDTVAPAVTIDSISDDTGAQANDFIT----- 1614
QY 983 NQVNTLLKGENLNKTDKNGTVTFGINTTSLGKAGKSTLNDGGLSIKNPTSEQIQVG 1042
Db 1615 -NDDTLVFNG-----TAEADSTVVVSLD--GIEIGTVTANGAGETWLDYGT-----LL 1660
QY 1043 ADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTCTNGSLDKSPHLSKOGINA--GKKI 1100
Db 1661 ADGDYELSVTATNPTGNSATATQTVVDTTAPTVTINAI-----AVDDIINAVEAGSPV 1714
QY 1101 -----TNIOGSEIAQ-----NSHDVATGKIKYDLKTELEN--KISSTAKTAQNSLHEFS 1147
Db 1715 AVSGTTTGVEDGVVTTIDGNTYATVTGNWTFNIPVADIANFEATEEVA-----T 1768
QY 1148 VADQGN-----NFTVSN--PSSSDTSTKSDVITF--AGE-----NGITTKVNGK-VV 1191
Db 1769 VSDLAGNPATPRNITVDTTAPTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVV 1828
QY 1192 RVGIDQTKGLTTPKLTVCNNNGKGI-VIDSONGQNTITGLSNTLIANTNDKGSVRTTEQG 1250

Db 1829 TVTID--GNTYATVTGNWTFNIPVADIANFEAT-----EEVVATVSDLAGNPATPR 1881
QY 1251 NIIRDEKTRAASIVDLSAGFNLQNGEAVDFYSTVDTFNFAFGNA---TTAKVTYDDT 1307
Db 1882 NI--TVDTVAPAVTIDSISDDTGAQAN-----DFITNDDTLVF-NGTAEADSTVVVSLDGI 1934
QY 1308 ---SKTSKVVDVNVDDPTTIEVKDKKLGKVTTLTSTCTGANKFALSNOATGDLALVRASD 1364
Db 1935 EIGTVTANGAGETWLDYGTLLADGDYELSVTATNPTGNSAT-----ATQTVVDTTA 1987
QY 1365 IVAHLNTLSGD--IQAKAGASOANNSAGYVDADGNKVLVYDSTDNKYQAKNDGTV----- 1417
Db 1988 PTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTTIDGNTYATVTGNWTFNI 2047
QY 1418 -----DKTKEV-----AKDKLVAQAQTPDGTTLAQMNVKSVINKEQVN- 1454
Db 2048 PVADIANFEATEEVVATVSDLAGNPATPRNITVDTTAPTVTINIAVDDIINAVEAGS 2107
QY 1455 -----DANKKQGINEDNAPVKLEKA--ASDNKTKNAAVTVGDL 1491
Db 2108 PVAVSGTTTGVEDGVVTTIDGNTYATVTGNWTFNIPVADIANFEATEEVVATVSD 2167
QY 1492 NAVAQTPLTFAAGDTCTAKKLGELT IKGGOTDTNK-----LTDNNIGVWAGT--DGFTV 1544
Db 2168 AGNPATPAT-----RNITVTVAPAVTTIDSISDDTGAQANDFITNDDTLVFNCTAEADSTV 2223
QY 1545 KLAKDLTLNLSNAGGT--KIDDKGVSVFVSSGQ-----AKA 1579
Db 2224 VVSLDGI EIGTVTANGAGETWLDYGTLLADGDYELSVTATNPTGNSATATQTVVDTTA 2283
QY 1580 NTPVLSA-----NGLDLGKV-LSNVGKTPKDPDAANVQQLNEVRNLLGLGNAGNDNA 1631
Db 2284 PTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTTIDGNTYATVTGNWTFNI 2343
QY 1632 DGNQV-----NIADIKKDPNSGSSSNRTV-----IKAGT 1660
Db 2344 PVADIANFEATEEVVATVSDLAGNPATPRNITVDTTAPTVTINIAVDDIINAVEAGS 2403
QY 1661 VLGKGNNDTEKATGGTQVG-----VDKGN-----ANGDLSNVW 1696
Db 2404 PVAVSG-----TTTGVEDGVVTTIDGNTYATVTGNWTFNIPVADIANFEATEEV 2456
QY 1697 VKTQDGSKKALLATVNAAGQTNLTNN---PAEAIDRINEQ--GIRFFHVND----- 1744
Db 2457 VATVSD-----LAGNPATPRNITVDTVAPAVIDSISDDTGAQANDFITNDDTLVFN 2510
QY 1745 GNOE---PVVQGRNGID-----SSAGKHSVAIGFQAKADGAEAAVAIGRQTOAGN----- 1791
Db 2511 GTAEADSTVVVSLDGI EIGTVTANGAGETWLDYGTCTLLADGDYELSVTATNPTGNSATAT 2570
QY 1792 QSTAIGDNAQATGDQSIAI-----CTGNVAVGKHSGA-IGDPSTVKAD-NSVS--- 1837
Db 2571 QTVVDTTAPTVTINIAVDDIINAVEAGSPVAVSGTTTGVEDGVVTTIDGNTYATV 2630
QY 1838 VGN-----NNQFTD---ATQTDVFG-----VGNNTVTES----- 1864
Db 2631 TGNWTFNIPVADIANFEATEEVVATVSDLAGNPATPRNITVDTTAPTVTINIAVDD 2690
QY 1865 --NSVALGNSAISAGTHAGTQAKKSDGAGTITTTAG--ATGTVKGFAGQTAGVAVSVA 1920
Db 2691 IINAVEAGSPVAVS-GTTTGV-----DGVVTVTIDGNTYATVTGNWTFNIPVADIA- 2744
QY 1921 SGNERIQNVAAGEVATSTSDAVNGSOLYKATQSI-----ANATNELD 1963
Db 2745 -----NFEATEEVVATVSD-LAGNPATPRNITVDTVAPAVERNELD 2784

RESULT 10

C48399

ABC-type transport protein ydba.2 - Escherichia coli

C:Species: Escherichia coli

C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999

C:Accession: C48399; D64891; H64891
R:Moszer, I.; Glaser, P.; Danchin, A.
Biochimie 73, 1361-1374, 1991
A:Title: Multiple IS insertion sequences near the replication terminus in *Escherichia coli*
A:Reference number: A48399; MUID:92190338
A:Accession: C48399
A:Molecule type: DNA
A:Residues: 464-2020 <MOS>
A:Cross-references: GB:D85081; NID:g3041754
A:Experimental source: strain K-12
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:88089)
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A6720; MUID:97426617
A:Accession: D64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-839, 'LDLPLYFQTSVIT' <BLA1>
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74483.1; PID:g1787665;
A:Experimental source: strain K-12, substrain MG1655
A:Accession: H64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'M', 915-2020 <BLA2>
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74487.1; PID:g1787672;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ydbA_2
A:Start codon: GTG

Query Match	5.0%;	Score 513.5;	DB 2:	Length 2020;
Best Local Similarity	21.1%;	Pred. No. 2.6e-12;		
Matches	426;	Conservative 266;	Mismatches 786;	Indels 539; Gaps
QY	307	DPKFAQNTNTKAGPLSTIGSINSIRKKLIINCA--GVNKTDVAVNVAQLAEVVKWA-----	357	
Db	159	DSVFTYTENAD-GTISLQDSNGRKATINLWQIDEANTVALEGVSDAGATKWOYNHNGEL	217	
QY	358	-----KERRITFQGGD-----NSTDVRIGLDNTLITIKGGAETNALDNNIGV--	399	
Db	218	VITGDNATVNNKKTIVDQKDSGTETENGNGKVIQDGLDVSGGHGIDITCDSATVDN	277	
QY	400	-----VKEADNSGLKVKLAKTLNLTENVNTTLNATTIVKVVSSTTAELLSDSLTFTQ	454	
Db	278	KGTMTVTDPSMGIDIGDKAIVN-NEGESTIINGGTGTQI-NGDDATANNNGKTTVDGK	335	
QY	455	PNTGSO--SSTSTVY-----GVNGVKFTNNAETTAAGITRIT-RDKTGFARDGV	502	
Db	336	DSGTETENGNGKVIQDGLDVSGGHGIDITGDSATVNKGTMTVTDPSIGIQVDGD-	394	
QY	503	DEKQAPYLDKKQLKVGSVAITID-----NGIDA--GNKKISNLAKGS-----SANDAV	548	
Db	395	--QAVVNN-----GESAITNGGTGTQINGDDATANNNGKTTVDCKDSTGTETIAGNCK	446	
QY	549	TIEQLKAAKPTLNAAGISVTPTEISVDAKSGNWTPPTYNIGVKTTTELNSDGTSKFSVK	608	
Db	447	VIQD--GDLDVSGGGHGIDITGDSATVNKGTMTVTDPSIGITQI-----DG--DQAI	497	
QY	609	GSGTNNSLVTAETHLASYLNEPVNRTADSAQSFVTKBEDDDANAITVAKDITKNACAVSI	668	
Db	498	NEG--ESTIINGGTGTQINGNDATANNNGKT-IVDQKDSGT-----TKIAGNIGI	544	
QY	669	LKLKGKNGLTAVATKKDGTFTFGLSQDSGLTIGKSTLNNBG-LTVKDTNE-QIQVGANGIK	726	
Db	545	VNLDDG--SLTIVTGAHGVENIG--DNG-----TVNNKGDIVVSDTSGIVGLNCEGAT	593	
QY	727	FTN-----VNGSNPGTGTANTARITPRDKIGFAGSDGAVDTNPKYLDQDKLVGNVKYINTG	783	
Db	594	VSNTGDNVNSNEATGFSITP--NSGRVSLAGS-----MVGQFDS-TGVDL-	635	

Qy	784	NAGKATIGLSPILPSI-----ADQSSRNIELGNTIQDKDKS---NAASINDILNT	831
Db	636	NGNNSVTLAAKDLKVVQKATGINVSGDANTVNI-TGNVLWDKDKTADNAAYFFDPSP	694
Qy	832	GFNLKNNNPIDFVSYDIDVFANGNATTAITYHTDANKTSKVYVDVNVDDTTIHLTGTD	891
Db	695	GINVYGDNNVTLDGKLTUVVSDEVTSROSNLFDGSAEKTSGLV-VTGDGNTVNMNG--	750
Qy	892	DNKKLGKVTTKLNTKSANGNTAFNWNSSDEDLVNAKDIKENLN---TLAKEIHTTKG	948
Db	751	-----GLELICEKNALADGSQVTSLTGTYSYTVIUVSGESSVYLANGDTTISGEF--PLG	803
Qy	949	TA-----DTAL-----QTFVYKKYDE-NNNADDANAITVQKNANNQVNTLTKGEN-	994
Db	804	FAGVIRVODKALLEIGSGATLMQODIDSPEHHGTRTPELTVADSCA-KIVNKGTVETQNL	862
Qy	995	GLNKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSIIKPTGTSEQIOVGADGVKFAKVVNN	1054
Db	863	GFAFVTGEN-----TGTNGSGISLLONG---KDPAPSPVILLATNG---GSATNA	9074
Qy	1055	GVVGCAGIDGTRITRDEIGF---TCTNGSLDKSPHLSKDGINA-GGKKTITNIOQSEIA	1109
Db	908	GTI-----TGKYTEQHSVENKYSTGTSNSFTFNDVDSITGLVLAQSNTIINTDSGLID	961
Qy	1110	QNSHDVATGKGIYDLKTELENKISSAKTAQNSHEFSVADEQGNFTVSNPYSYDTSK	1169
Db	962	LYGRGSGVGLAIAD-----STAENQCKITLDSMWVDANDTTA	998
Qy	1170	TSDVITFAGENGITTKVNGV-VRVGIDOTKGLTTPKTLTVGNNGKGVIVDSQNGQNTIT	1228
Db	999	MRTIAS-----NSAIDFGTGVGVGTDTSYG-AGKNATAINQLGGVITTYNAGAGMAAY	1050
Qy	1229	GLSNTLANVTNDKSGVRYTTEOGNIKDEKTRAASTVDVSAGFNLOQNGEAVDFVSTYD	1288
Db	1051	GASNTVIN---OGTINLEKNGV---DLSLAANTIV-----GHAVYEHGFAIN-----D	1093
Qy	1289	TVNPFADGNATTAKVTYDDTSKTSKVYDV-----NVDDTT-----	1323
Db	1094	QTGVININVGTCGAFYNDGTGTI-VNYGTCITFGVCQSGNEYNTDDFTSLIYTGDDTIT	1152
Qy	1324	-----IEVKDKKLG--VKTTTLTSTGTGANKFALSQAOTG-----DALV	1360
Db	1153	RSGETVTLNKSAAVTDKLAGVNVNSGTLSDGQITVYSSGELLENTSGTIINLVKLDKGAVI	1212
Qy	1361	KASDIVAHLNLTSGDIQTAKG--ASQANNSAGYVDADGKNKVIYDSTDNKYYQ---AKND	1414
Db	1213	KNAGVWNTNVYDSGGLINNAGEMTAQITWNAG---ADSLVNNTGCTINKIVQNGAGVFNNS	1269
Qy	1415	GTVDKTEVAKDKLVQAQTPDGTL---QOMNVKSVINKE-----QVNDAN	1457
Db	1270	GSV--TGRMMSAGGVFNQOT-DGAIIRGAALTCTAVANNEGTVNLGSSSEGNNTGMLEVN	1326
Qy	1458	KKQGINEDNAPVVGLEKAA-----SDNKTNAAVTV-----	1488
Db	1327	NNSAFNNRGEFILDNDKNAHVHINQSGTLYNTGHMIMISSHNGAYNMWGGNGREINDGTI	1386
Qy	1489	-----GDLNAVAQTP---LTFAGDTGTTAKKIGELTTLTKGGOTDTNKLTD	1530
Db	1387	DVSXKSLVSNANNQONQAFNNQDNGVINFDHDSASAVK-----VTHSNFTAQ	1435
Qy	1531	NN-IGVVAGTDGFTVKLAKD--LTNLNLSVNWAGGTRIDDKGVSFVDSQQAQANTPVLSAN	1587
Db	1436	NDGIMNISGTVGAVAMEGDKNAQLVNNGTINLCTAGTTDTGTGIMGOLDANATADA-VIENN	1494
Qy	1588	G-----LDLGKGVISN-----VKGK-TKDDTDAANYQQLNEVRNLLGL	1623
Db	1495	GTINIFANDSFAFSVLGTGVHVNVNCTVVIADGVGTSGGLIKOGDSINVEGNN-----	1546
Qy	1624	CNAGNDN-ADGNQVNTADIKK-----DPNCGSSSNRTVTKAGT-VLGGKGNNDTE	1671
Db	1547	GNNNGSEVHYGDYTLDPDVPKPNVTSVTSGSDEAGSGMNNLNGYVVGTFNNGSAGKLKVN	1606

Db	1589	AGIILDNEGTLANGAQLLATPCTLLLRQLGVNDNSGGEISSDRAFTLTTSALTNOQGR	1648
Qy	310	-----FOATNNTKAGPLSIGNSTKRKII-----NVGAGVNTKDAV--NVAQL---350	
Db	1649	LRSGVLTLRTAQALDNLSEGLVS-GTGGGLDIRALALDNRSGTSGKGAVIDIDVSRLEND	1707
Qy	351	-----EAVYKWKAKER-----RITFGOD-----DNS370	
Db	1708	DGDLSEGRKLKLTAEANSVRGRTAARGDLHASVTAFNQAGGELSGALMLEADSLNDR	1767
Qy	371	TDVKIGIDNTLITIRKGAFETNALTDNNIG-----VVKEDN-----SGLVK411	
Db	1768	SGGLVSADGNLTVSA-----RRIDNRAGEIASPQVVTLDVAEQIDNRKGAIGDSGLURLA	1822
Qy	412	LAKTLNN-----LTEVNTTILNATITVKVS--SSSTTAEILSDSLTFPTQNTQSQTSK	464
Db	1823	APRVLNQDGVLASRGURLNGAELFNCGNGLLSQQSIDVILGV-----LGNQAGSL	1876
Qy	465	TVYVG---VNGVKFTNNAETTAAGITT-----RTRDKIGFARGDGVDERQAP	508
Db	1877	SSQGRLSVKSGRLDNQGGAVSSAGTSLSSOGALNNQGRVVTDAGAVLRASLSDNSQGG	1936
Qy	509	YLDKK---OLKVGSAIITIDNGI--DAGNKISNLAKSSANDAVTIEQLKAAKPTLNA-	562
Db	1937	IVSAKGAABEIRTGSLNNSQKGGIGSGAGIALVADILV-DNSQNGRIT-----AKGAIDAN	1989
Qy	563	-----GAGISVTPTEISVDAKSN-----VTAP-----TYNIGVKTTELNSDGTSD	603
Db	1990	LKGLDQOQSGRLVSDTAALDLRGCELLVNRQAQGLIATPGALLRLQGVVNDNSGGGEISSD	2049
Qy	604	K-FSVKSGSTNN---SLVTAEHLASLYELNVRTADSALQ-----SPTVKEEDDDANA	652
Db	2050	RSFTLAATAALSNRGRVITSGDSLTL---RIAQALDNLQGVLSAGGLDVAALVFEDNISG	2106
Qy	653	ITVAKDTT-----KNAGAV-----SILKLGKNGLTVATKKGTGVTFG-LS	692
Db	2107	IVASKGDTHIGVNRLENEAGRVVSEGAOLDLTAKQVSSAKGR---IAAKGDLQVTVGTLE	2162
Qy	693	QDSGLTIGKSTLNNDLTVKQDNEQIQVGANGIKFTNVNNSNPGTGIANTARI-----	745
Db	2163	QQGGELASQGTLLDADSLDNRNGGLVSADGGVTAEARQIDNRGGEISSVAKVALAVREQ	2222
Qy	746	-----TRDKIGFAGSD-----756	
Db	2223	LNRGKGVIIGDSELSITVQRLNQAQGVLASRDGLHDGAELLNGDGLLSQRLVDVTL	2282
Qy	757	-GAVDT--NKPYLDQDKLQV-----GNVKITNTG--INAGKKAITGLSPT	796
Db	2283	SGALDNOQSGALVSEESLTVKADQVNNQAGTFSFSSAGLSLTVSRGELNNOGQVLTAGAT	2342
Qy	797	LPSIADOSSNIELNGTIODK-----DKSNAASINDILNTGFNLKNNNNPIDFVS	846
Db	2343	LNSTGFDNSR-----AGLVSAGKAVATRTGALNNSQKSIGG--NTGVTL-----	2385
Qy	847	TYDITVDPANGN---ATTATVTHDITANKYSK-----VYDVNVDDTTHLTCTDD	892
Db	2386	VAGLVD--NGREGRISTKGTLDANLKGLOQGGGSLVGERGVTLDLN-----GGTLD	2435
Qy	893	NKKGIVKTK-----LNKTSANG-----NTATNFNVNSDEDALYNAKD	931
Db	2436	NHDGLVSTPGALLRLQGVNDNSVSGEISSDRAFTLTAANTLNOGGRLLSSEAL--TLR	2493
Qy	932	IAENL-NTLAKEIHTTKGTA-----DTALQTFVK-----KVDENNADANAIVTG--	977
Db	2494	IAKTLNLSKGQVLTADGLAIESQVLDNRAGTIGSKGDARISVTSLDNAEQGSLVSEGRL	2553
Qy	978	-----QKNANNO-----VNITLTKG-----ENGLNIKTDKNGTTFGINTTSG	1015
Db	2554	ELVADQVSNQGRRIARGVLEAAVGTLLQOQGGELVSQGSLLDLRAD---TLD---NSQSG	2607
Qy	1016	LKAGKSTLNDGGLSIIK-----NPTG-----SBQIQVADGVKFKFKAQVNNNG--	1064
Db	2608	LJAA-----NGGIAEAROVNDRNAGEISSTSQKAVNAR---EQLDNRNGKGVK---DSG	2655

Qy	1065	TRIRREIDGFTGTNGSLDKSKPHLS- KGINAGGKKTTINTQSBEI -AQNSHDAVTGGKIY 1122
Db	2656	LRLTVQRL-----LNOAKGVLAGRDGLSDGELFNGDGRGLDSQNSLSVSLAGVLD 2707
Qy	1123	DLKTELENKISSTAKTAQ-----NSLHEFSVADEOQNNFT-----VSNPYSSY 1165
Db	2708	NOGQALVSEGSLTARAARLDNRGFTFSAGALALITSQAVLDNOQGRLLSDAGVTLKGASL 2767
Qy	1166	DTSKTSDVITFAGENGITTKV-----NKGVVRV-----GIDOTKGLITTPK 1205
Db	2768	DNSR- SGVISAQAVDIRTGVLDSNRNGIGISNAGITLVAARLONQOGRVSAKGLLDAN 2826
Qy	1206	LTGVNNNGKIVIDSONGQNTITGLSNTLAN-----VTNPKGSVRTEQGNII 1253
Db	2827	LKGLDQRGGGVIV-SETGV-TLDLNGGTLVNRDGLIATPGALLRLRLQJAVDNCAGGETS 2884
Qy	1254	KDEBKTRAASIVDVLSAGFNLGN-----GEAVDFVSTYDTVFNADGNATTAKVTYDD 1306
Db	2885	SDRAFTLAAASLD- NRGRELIGADSLTILRQAOLD- NSLAGVISGAAG- LDTAAARLDN 2940
Qy	1307	TSK-----TSKVYDVNVND-----DFTTIE- VKDKKLGVKVTTLTSTGTG-----ANKFA 1349
Db	2941	SAKGTILASRAGIDRLRVDGALDNHAEGTVSGARLTLASASLDSNCKGLLSGNAGLSVATGA 3000
Qy	1350	LSNOATGDAL-----VKASDIVAHLNTLSGDIQTAKGASOANNSAGYVDADGNKVI--- 1400
Db	3001	LDNAEGQLISOGVLDVSSADLDNRGGALSCKQSLRLISAANLDRGGLLTSDGELELTAG 3060
Qy	1401	-YDSTONKYQAKND-----GTVDKTEVAKDKLVAQAQPDGTFLAQMNVKSV 1447
Db	3061	RVDSADGGEISARGDLRLTVERLVORQGRLTGERCVSLDLRGGDLNNOGGLISARGPLSI 3120
Qy	1448	INKEQWANDANKQGINEDNAFVKGLEKAAS--DNKTKNAAVTVGDLNAVAQT----- 1497
Db	3121	---ERLVLDNRQG--GEIYSQGFELLARRIDNQGQRIISAGKRLDADALGNAGAGL 3175
Qy	1498	---PLTFAGDT-----GTTAKLGE-----TLTIKGQDTNKLNTDN 1531
Db	3176	LSWQHGTVTGGSLDNSAGGTLSSKDGELATISGALDNHGGQALVSKGAQRIDAASLDN 3235
Qy	1532	NIGVAGTDGFTVKLA-----KDLTNLSVNNAGTKIDD-----K 1566
Db	3236	AQGVSGESDVTLSIAGKLDNGGGGLVSAQRALSFERDDTLN--NAGG- RINGSSILLK 3292
Qy	1567	GVSEVDSSGQAKANTPVL SANGLD- LGKVISNVGKGTDTDAANVOQLNEVRNLIGLG 1624
Db	3293	GASLDSNDGO-----LISQGRLDAILGALV-----AGAARLASGDLL-LR 3334
Qy	1625	NAGDNADGNQVN- IADTIKDPNSGSSNRTVIKAGTVL--GGKGNNDTEKL-----ATG 1676
Db	3335	SASVDNRGKLLVSOGLELISAGSLDNSASGTLASQADMSLRIGGGALRNQODGLITFSQAG 3394
Qy	1677	GIQVGDKDNANGDLSNYYVTKRDGSKKALLATYNAAGOTNYLTNNPAEADIRINEQG 1736
Db	3395	ALEVQAGSLDNROGTL-----QAQGDNRILRIGGALDNQAGRILDSRAGN----- 3437
Qy	1737	IRPFHVNDGNQEPVQGRNGIDSSAGKHSVAIGFAQKADG-----EAAVAIGROTQA 1789
Db	3438	---LDLQSGSLD--NGAGGVLSNAKGMKLKLTGLFDNSAGVTQAOLEIRAGQGVNRNQ 3491
Qy	1790	GNOSIAIGDNAOATGD-----QSIATGTGNV-ACK----- 1819
Db	3492	GHLSALGGDNRIVTADFDNQGGGLIYASGLSLDQQRFLNQGAAGQCGKVCAGRIDFSLA 3551
Qy	1820	-----HSG-----AIGDPSTVK-----ADNSYSV-GNNNQFT 1845
Db	3552	GALANRFGQLESESELHRAAIDNSGGSRLRGRSGSTRLVAGDLNNAVGVLESANQDL 3611
Qy	1846	DATQTDVFGVGNNTVTSNVSVALGSNAISAG-----THACTQAKKSDGTAGTTTATAG 1901
Db	3612	DLOGLSLANAGRIILHTNGTGLDSGVIRAGGELTINGLLDIRASEFTNSSVLOAGRLL 3671

Qy	1902	GTVKGAGQATGAVGASGAERRIQNAAGEVSAITSDAVNGS-----QLYKATQSTANA	1958
Db	3672	NLDICTFQTAEGKILLAVQSFTGR-----GCDWSNDCLLASNGSLRLELSGGYRGNGRA	3725
Qy	1959	TNELDHRTHQNEKANAGISSAMAMASMPQAYIPGRSNVTGGIATHNGQGAQAVAVGLSKLS	2018
Db	3726	TSLGDFAL-----NAASLDLGNASLA-----GGANVTLGAGLLVNRGRIT	3767
Qy	2019	DNGQWVFKINGSADTQGHVGAAGVG	2044
Db	3768	AAGDLV-----ASAASLNNGYTLGGGG	3789
RESULT 12			
T31102			
filamentous hemagglutinin 1 - Haemophilus ducreyi			
C:Species: Haemophilus ducreyi			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999			
C:Accession: T31102			
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.			
J. Bacteriol. 180, 6013-6022, 1998			
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.			
A:Reference number: Z20984; MUID:99030326			
A:Accession: T31102			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-4152 <NAR>			
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1			
C:Genetics:			
A:Gene: lspA1			
Query Match 4.9%; Score 506; DB 2; Length 4152;			
Best Local Similarity 20.2%; Pred. No. 1.3e-11;			
Matches 494; Conservative 342; Mismatches 864; Indels 744; Gaps 126;			
Qy	2	NHIYKVPNKATGTFMAVAYAKSHSTGGGSCATGQVGS-----VCTLSPARIAAL	52
Db	3	NKRYKLIFSKVKNCILVPVAENIKSAGNSGSSSKNTAEQDEEPDSLACSL	55
Qy	53	AVLVIGATLGSAYAKKDTKHIAIGEONQPRSGTAKADGDRAIAIGENAAOGGOAIA	112
Db	56	-----PLSSS-----IHLGLNH-----	68
Qy	113	IGSSNKTWNGSLDKI-----GTDATQGESI--AIGGDVKASGDASTAIGSDDLHLDD	164
Db	69	-----SPLKVPKGSLSVLLSLMPATPLLAQONYAALANGKVYVDSQHS-----STRIYEQ	120
Qy	165	HGNPKHPKGTTLNDLINGHAVLKEIRSKDNVRYRTTASGHASTAVGAMSQAQGFHN	224
Db	121	KTNDNSKDGIVVVEIAN-----PEVDGVSND-----RFKEFNIPNSAV-----FNN	161
Qy	225	AFGTRATAKSAYSLAVGLAATAEQGSTIATGSDATSSSLGAIALCAGTRAQLOGSIALG	284
Db	162	S-----RTESTSOLVG--KLHAKIQLQKEAKLLNQ	190
Qy	285	GSVVTQSDNNRPAYPNTQALDPKFOATNNTKAGPLSIGNSIKRKILNY--GAGVNKTD	343
Db	191	-----VTGDHES-----NIQ-----GAEVACKKADLLIVNPNGITLNGVK	226
Qy	344	AVN-----VAQLEAVVYKWKARRITFQGGDNDSTVKIGLDNTLTIKGGAETNALDNNIGV	399
Db	227	TINTDRFVVVSTDIIPHRENGLLSVRNG-----KVID-----KGGVATNGLSHFEV-V	274
Qy	400	VKEADNSGLKVKLAKTLN--NLTEVNTT-----TLNATTYVKVGSSSSTTAELLSDSLTFT	453
Db	275	ARNIDQKG-KITYAKTENQKSVNPNITFAAGSLNLYNLTREATPISSGTSRTSDTPAIS	333
Qy	454	QPNTGQSQSTSKTVYGVN-----GVKFTNNAETTAICTTTRTRDKIGFAR	498
Db	334	ADSAGS-----MYGSNLIKFFVYDTKGACVKKHKGIIFSEN-DINKIMDGGNASKULAKYAK	386
Qy	499	DGDVDEQAPYLDKKQLKVGSVAITIDNGIDAGNKKISNLAKGSSANDAVTIEOLKAAKP	558

Db 387 DILAKDIELTEKQLOANNKIIL-----NSTGKINLRNASEVSDNVDNVNSENALENA 442
Qy 559 TLNAGACISVTPPEISVDKSGNVTAPTNYIGVKTTELSDGTS---DKFSVKGSGTNNNS 615
Db 443 SMSANS--LDVIVTKIEVN--RSSKVSAGTANI--KASNIITLDGSSVANKIIL--NVTNNA 496
Qy 616 LVTAEHLAS-----YLN-----EYNRADSALOSFTVKEE-----DDDDANAIT 654
Db 497 TLNNSKLSAKDMELNVTHNITLNNTSKLSAOKANIKTENLTNGEASLVAEKLINAI-- 555
Qy 655 VADTTNAGAVSILKLGKNGLT--VANK-----KDGVTTEGLSDGSLTI--GKSTLNND 707
Db 556 ---DKITNNGTIA-----GUTANITTKALENRDNALILA--HQLNFTVNGSHVYVNGK 603
Qy 708 GLTVKDTNEOIQVGANGIKET-----NVAGSN--POTGIANTARITRDKI 750
Db 604 DIVSKD--KAIVTFNSNDSFTSNGSKLVDAQNLTNVNMFNITOGSEIILHGNVTLNKA 661
Qy 751 GFAGSDGAVDTNK-----PYLDQDKLVQV--NWKI--TWGINAGSKAITGLSPTLPSI 800
Db 662 GNFTNSGNLTMTKELNISIESEFINAGNLTTGKNLEVHSNNTVKNDOGLVSIENLMSISK 721
Qy 801 ADQSSRNIEGL-----NTIQDKDKSNAASINDILNTG-----FNLKN 837
Db 722 TDFTNNGTLGLEALKTASGNGFTNASNGSLASNKSLDIYGNFTNNGTIESVKSLNITN 781
Qy 838 N-----NNPIDVPSYDIIVDF--ANGNATV-----ATVTHDTANKTSKVYDVYVNDVTTIHLT 888
Db 782 NYTFENNTAT--IKSYGLNITISOGNFTNDSNGTVMSHDLNITSQ-----ANIKNKLLAG 835
Qy 889 GTDDNKKLGVTTKLNTKSANGNTA-----TNEVNVSSDEDALVNAKDI---AENLNTLA 940
Db 836 GOGGLN-----LTAKGNTINDSNSTALVLSNNDINLANNKYNVIGEIYSQAGNISVEA 890
Qy 941 KEIH-----TTK--GTADTALOTFTVKKVDENNADANAITVQKNNANNQVNTL 988
Db 891 KLLHNDVKLSGNTITTKSGNA-----TVKNTSIOGGLHDANSIRVE-----L 934
Qy 989 TLKG-----ENGLNKTDKNGTIVTGTINT-----SGLKAGKST-----LNDGGLSIKNPT 1034
Db 935 TLNGKFAOLDN--QULKVALRKIYAGSNLTTPFAKEGEKEQKSTAAQKIIRGTINVRNKL 992
Qy 1035 ---GS-----EJOIQVG-----ADGVKFAKVNNGVVGAGIDGTTT 1068
Db 993 EYGSNVVDENNMRSMQVNLYEKIFNGDNPIITLLKNGVTFKAFDSNRRRASNDG----- 1047
Qy 1069 RDEIGFTGNSLDSKPHLSKDGINAGGKITTIOGGEIAQNSHDVATGKI---YDL- 1124
Db 1048 -----EGTNKKTFDNVAHLEAF-----SGYSNGNDHRASDDGHVKSPIYLL 1090
Qy 1125 -----KTELENKISSTAKTAQNSLH--EFSVADEQGNFTVSNPYSSY----- 1165
Db 1091 VLAQAVNTEGENYL---KTALQHIFGPNWMDLTNTTNDTINDKNQLKWKFKNN 1146
Qy 1166 -DTSKTSVDITPAGENGI--TTKVNKGVVRVGDQTKGLTTPKLTVGNNGKGVIDS--- 1220
Db 1147 GENNHSINLIYPADEGVKAKIFAGVLNNG---TNGVEDKVVQELNDKAKKEYEDKFAK 1203
Qy 1221 -----ONGONTITGLSNTLANVTNDK--GSVRTTEQGNIIKDE----- 1256
Db 1204 KFOGRKPSRFQNGEFDWAG--DWAKEGNEYSYKSETEEKYNGIKKEHTVNIKGKHEIKVP 1260
Qy 1257 -----DKTRASIV-----DVSAGFNLOONGEAVDFVSTY 1287
Db 1261 TVSFENLNNHQQSDGDKSIISSELLAQPIIYVAKADVDPDVPRAONKRAVEDGGLY 1320
Qy 1288 DT-----VNEADGNATAKV-----TYDDTSKTSKVYVYDVNVDDTTI-----EVKDKKLG 1332
Db 1321 RTRLSYLN--QNYNLGAKYFFNQLDTEDDKLGKIGRIGNYEFHQILTRLEIKVADNHLT 1378
Qy 1333 VKTTTLTSTGTGANKFAL-----SNQATGDAVYK-----ASDIVAHLNT- 1371

Db 1379 LK-----HGLHDIALVKKLIDSASIOAKDLNLNKVGEALTKEQKDNLKEDIVWYVTKTE 1430
Qy 1372 -----LSGDIQTAKGASQANNAGVYDADGKNKVIYDSTDNKNYYQAKNDG 1415
Db 1431 VNAOEVLVPOVYLAKQTIIEVEKORGVGTCQIRAGIIV-----KVDDVRNTG 1478
Qy 1416 TVD--KTKEVAKDKLVAQAOTPDGTLAQMNVKVINKEQVND-----ANKKQ 1460
Db 1479 TIAGVAVGLEAKNKL-----KNTGDILSORLSKLGVKKGLESTGCVTVYVDETGAIVKVRK 1533
Qy 1461 GINEDNAFV-----KGLEKAASDNKTKNAAVTVGDLANA--VAOTPLTTF-----AGD 1504
Db 1534 IKSEGHYLETDKDNVDLTASELKGNTGQIKAKDLNLDIYETSYKYKELFKNGGE 1593
Qy 1505 TG-----TTAKKLGET-----LTIKG--GQPTDNKLTNNIGVAVGDTGFTVVKLAKD 1549
Db 1594 IGRVTOQSQAQSVGTDAFPHLHLSLEGDVNQTSNLSKANRTTGVYVGK--DFNTKAGKD 1651
Qy 1550 L--TNLNSVNAG-----GTKIDDKGV--SFVSSSGOAKANTPVLSANGLDLG 1592
Db 1652 LFHRQIDVTVTGTVYSASASGGGOSAGISLTDQGVETTYTNKTATAGANADV----- 1702
Qy 1593 GKVISNVCKGVTQDAAVQQLNEVRNLLG-----LGNAGNDNADGNQ--VNI-----ADIKK 1643
Db 1703 ---TNFMKRTREYETSLTHRSEENALSGELYVMGRADIGGVINDRQVDEVIKTPPEIAA 1758
Qy 1644 DPNSSSSNRVTIVKAGTVLGGKGNNDTEKLGATGGIOGVCDKGNANGDLSNVWVKTKODG 1703
Db 1759 EQKAAEEAKKAEVK-----ENEASETAAKETE--EAENDNVAEKDKTKPKFKKLT- 1807
Qy 1704 SKALLATVNAAGOTNLTNNPAEAIDR-----INEQIRFEHVNDGNQEPVVGNGRND 1758
Db 1808 --EETAAAPETGCEFFPAAYKARBEEDRKKGFTLSAEQIESTKARDEXETTYVYELKVGVG 1865
Qy 1759 SSASGKHSVAIGFOKA-----DGEAAVAIGRQTCAGNQSTAIAGDNAQATCDQ 1806
Db 1866 ABEAESAADAISNKARQIIDTQNGLKQDGTVAL-----QEASDVNLATGLAGSAKL 1921
Qy 1807 STAIGTGVNVVAGKHSAGIADPSTVKADNSYSVGNNNNQETDATQTDVFGVGNNTVTESNS 1866
Db 1922 KPELST---IEKSRGA-----SDGRSILGRLNLA--ARGGDI--TLNNVETTENS 1967
Qy 1867 VALGSNSAISAGTHAGTQAKSDGTAGTATGTTTATAGATGTVKGFAGQTAGVNAVSVGASAEER 1926
Db 1968 LSLKARDNVNVSGV--TEQKDESNQSOLKVTACASSCGGVAGGCSAG--VSAGVSGS--- 2022
Qy 1927 IQNVAGEVSATSDAVNGSQL--YKATQ--STANATNEILDHRTIQHNENKANAGISSAMAM 1983
Db 2023 -YNESNTESTSHTNSLIRKSLRVEAGKDFNLISSNVVDVH--LHL--DVKGDTNVVYSKQDS 2079
Qy 1984 ASMPQAVIPGRSMVTGGIATHNG---OGAVAVGLSKLSLSDNGQWV 2024
Db 2080 YSRKRGVNVSVSAGVGVSTAGARPNGSVGLGVSAENENSKIV 2123

RESULT 13

T34434
hypothetical protein K06A9.la - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Giesel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2232 -CEI>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.la
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.la

A: Map position: x
A: Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

```

Query Match      4.8%; Score 494.5; DB 2; Length 2232;
Best Local Similarity 18.0%; Pred No. 1.6e-11;
Matches 385; Conservative 326; Mismatches 958; Indels 467; Gaps 72;

QY 17  MAVA EYAKSHSTGGGCATQGVSVCTL-----SFARIAALAVLVIGATLGSAYAQKKD 71
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 LPIASSASSSPSAASSTTPWLSSSTIQSSSGTFPSSVASSPSTVGST-SGAASSSSYA 310
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 72  TKHTAICEQNP RSRGTAKADGRARIAIGENANAGQCAITAIGSN-----KTV 120
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 TVSTIAG-----STGSTITPVPGSSSTIGSTSPSASSSSSGTMTSTI 351
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 NGSLDKI-----GTDATGQESIAIGDVKASGDASIAIGSDLLHLLDQHGNPKPKGTLI 176
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 SGSTGTVTVVPGSSSTFASSTPTASSSSPGSTTVVAPGSS----- 393
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 NDLINGHAVLKEIRSSKDNVDKYRRTTASGHASTAVGAMSYAQGHFNAFTRATAKAS 236
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 394 -----TVGSSTPSASSSSS-CTMSTNSGSGTGTVTVPVSSSTF 431
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 SLAVGLAATREGQSTIAI--GSDAT--SSSLGAIAGAGTRAQLOGSIALQGGSVVTQSD 292
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 432 GSSTPIASSSSSGSTTVVAGSSSTYTGSSSTPSASSSAGTASTIGS---TGSTATIVP 487
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 293 NNSRPATVPTOALDPKFAQNTNNTKAGPLSITGNSIRKKIINCAGVGNKTDVAVVAQLEA 352
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 488 GSS--SSVGSSTQASPSPGTMTSGP--TGST-----VTVPGSGSTSPAPS----- 532
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 353 VVKWAKERRITFGDDNSTDVKIGLDNLTIKGAETNALTDNNIGVVVKEADNSGLKVKL 412
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 533 -----SSPNPSSSPASTGSTITISGSSSI-----IVSTVSGS----- 564
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 413 AKTLNNLTENVTTLNNATTVKVGSSSTTAELLSDLSTFTQPNGTQSOSTSKTVYGVNGV 472
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 565 --TVSGTGTSQSTLASSATATPGSSSTVPSSSPQSPQSPAPNTGSTITPSQT--SSQSP 620
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 473 KFTNNAETTAAGTTRITRDKIGFARDGDVDEKQAPYLDKKQKLVGSVAL-----RIDN 526
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 621 SPSPNPSSPTGSSQSTIIPEGSTASPGTSGSTP-----SVATEVTSQSTVPS 671
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 527 GIDAGNKKISNLAKGSSANDAVTTIEQLKAAPTLNAGAGISVTPTEISVDAKSGNVTAPT 586
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 672 GSSLGIGTSSNPSPSSLSPTSCMSTLTSEP-----SPSSQTQSSCAQSTLTTPS 721
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 587 YNIGVKTTENLSDGTSDFKFSVKSGGTNSLVTAEBLASYLNEVNRNRTADSAQSFTVKEED 646
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 722 PNPQSSTSSLES---STGATTSSGAGTMTSPSSSVGS-SQGSTSPAASSTSGEMT 777
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 647 DDANAITVAKDTTKNAGAVSILKLNKGLTVATKDGTVTFGLSQDSGLTICKSTLNN 706
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 778 SQGSTQTPGSSVSTSAAILTSTQDSVTNPGSTVTPRPSTVSGSTSSGSTTVGSTEAST 837
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 707 DGLTVKDTNEQIQVGANGIKFTNVNGSNPGTGIANTARITRDKIGCFAGSGCAVDNPKPYL 766
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 838 SGSSVASSSPAPTSQN-----PNPST-----SSGSSMITQSPYP 872
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 767 DQDKLQGVNKYITWGINAGGKAITGLSPTLPSTADQSSRNIELGNTIQDKDKSNAASIN 826
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 873 SQSTSPVSSSTPSPG--SPCTTUTSTSPS-----PSQSTTICST----- 910
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 827 DILMTGNLKNNNPIDFVSTYDIDVFANGNATVATVTHDTANKTSKVVDVNVDDTTIH 886
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 911 -----QGSTSP--GISTSEMTSQGSTQTPGSGTGTVPQPS-TVSDSTSGSGTGT 958
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 887 LTGTDNNKLGKVTTKLNKNTKNSANGNTATFNVNNSDDALVNAKDIAENLNTLAKEHTT 946
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 959 VGSTEGS-----SSPIPTSFQNTNPSTSGSSMSTQTP-----OSSOSTSPVESST 1005
Db      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 947 KGTADTALQTFVTKVVDENNADDANAITYGQKNANQVNTLT-----LKGGENLINI--- 998

```

Db	1006	GATSSSGPCTTITSI--SPSPSPSTIGSSOGSTSPVVSTISOGSTETPGSTGTWTKP	1063
Qy	999	-----KTDKNGTVTFGINTTSGLKACKSTLNDGGLSIKNPTGSEQIQVAGDGVKFAK--	1051
Db	1064	STVSGSASSGATMKGSTASSTSGSGSTSPNPSQSTSPSTSGATSPGSGGTLTISLP	1123
Qy	1052	--NNNGVVGAGIDGTTTRIDREIGFTGTNGSLDKKPHLSKDGINAGGKKITNIQSGBIA	1109
Db	1124	SPSQSSTIGSSOGSTSPVVSTTSGDMTSQGSTQIP-----GSGTSTVQTPSTG--	1171
Qy	1110	QNSHDVATGKKIYDL-KTELENKISHTAKTAQNSLHEFSVADEQGNNTTVSNPFYSYDTS	1168
Db	1172	--SGSTSTGEITSOGSTQTPRSSLSTSPASTSTQQSVSTNPSGS--TVTPQSTVTRGST	1227
Qy	1169	KTSDVITTFACENGITTKVNGKVVVRGID-----QPKGLTTLTKPLVGNNGNGKI	1216
Db	1228	SSGTVTGTGTEGSSSTGSSSATSLSSSPVPSTQSPPNPSTSGSGTP--TPNPSQSTSP	1285
Qy	1217	VIDSQNGQNTITGLSNTLANV-----TNDKGSVRT--TEOGNIILKDEKTRAA	1262
Db	1286	VVSTTTGEMTSHGSTQTPSTIGSTVQTPSTVSGSNSSGTVTIGSSEASTSGSFSKTPS	1345
Qy	1263	SIVDV-----LSAGFNLOQNGEAVDFVSTYDVNPA-----DGNATYAKVYDDTSKT	1310
Db	1346	SISVPSTSPITPSTFASSTSGSTISDVSVSTTSLAPLSLSPSTVPSPSTQSFSSTSEG	1405
Qy	1311	SKVVYDVNDVDTIEYKDKKLK--VKTPTLTSTGTCAKFKAL-----SNGATG	1356
Db	1406	SSKASSPVPSTQSTPTNPDTGESTSTLLSSTIGSTQHTTMSKASSGSTSPSTNSGTG	1465
Qy	1357	DALVKASDIVAHLNLTSG-----DIOTAKGSAQNNAGYVDADGNKVYDSTONKYYQA	1411
Db	1466	STVTMGSSSTSGVSTSSASTQPMQSTQMGSSAGSTVASSSTASPAASSTAPSTGT--MS	1524
Qy	1412	KNDGTVDKTEVAKDKLVAAQAPDGTQLAMNVKSVINKEQVNDANKKOGINENAFVKG	1471
Db	1525	TSSGTVGST--ISESSTTASQSTGSTVTVMGSSST-----SGVSTSSA--SS	1568
Qy	1472	LEKAAQDNKTKNAIVTGDNLNAVQPLAFAGDTGCTAKKLGELTILKGGQDTDNKLTDN	1531
Db	1569	TPQMSTSQSSAGSTVASSTAGLVSTSTVPSTG-----MGST-----	1608
Qy	1532	NIGVAGTDGFTVKLADLTNLNSVWAGGTKIDDKGVSPVDSGQAKANTPVLASGLDL	1591
Db	1609	-----SSGTVGSTISESSTTASASQSTGVSTVMGSSSTSGVSTSSASTQPMQSTSQGSSA	1664
Qy	1592	GKGVISVWGKTQDQDAANVQQLNEVRNLLGLGNAGNDNAGNQVNIADIKKDPNSGSS	1651
Db	1665	GSTVASST-----TGLVSTSTVPSTGTMGSTSSGTG-----STISESSTAAAS	1710
Qy	1652	NRTYIKAGTVLGKGNNDTEKLTATGGIQVQVDKGNANGDLNVNVKTKQDKGSKALLAT	1711
Db	1711	SQT---GSTVT--MGSSSTSGVSTSSASGQPMQSTSQG-----	1744
Qy	1712	YNAAGQTNV--LTNNPABADIRNEQGIREFHVNDGNOEPVQVQNRGIDSSAGKHSVAIG	1770
Db	1745	--SSAGSTVVGSTASPAAS-----STAPSTGTWSSSTSG	1777
Qy	1771	FQAKADGEAAVAGROTQAGNQSIAGDQNAQTGDQSTIAIGT--GNVVAGKHSGAIGD--	1826
Db	1778	TVGSTMQSQTAAASTTSHTGS--TVTLGSSSTSSNQMSTSQGSSVGVSTVASSTAGLVSTST	1836
Qy	1827	-PS---TVKADNSYSVGNNOFTDQTQDVFVGNNITVTVESNSVALGNSAISAGTHAG	1882
Db	1837	VPSSTGTMGSTSSGTSTISESSTTASASQSTGTVTVMGSSSTSGVSTSSASSTQPMQS	1896
Qy	1883	TQAKKSDGTAGTTTAG--ATGTVKCFAGOTAVGAVSVCASCAERRIQNVAAGEVSATST	1940
Db	1897	TSQSSAGSTVAGSTAGLVSTSTVPSTG--TMGSTSSGTVGSTTISESSTAASTSSOTGS	1954
Qy	1941	DAVNGSOLYKATOSIANATNELDHR-IHONENKANAGISAMAMAMPQAY-IPGRSMVT	1998

Thu Sep 13 14:18:58 2001

Db 1955 TVTIGS-----TSGTNPSPRSLQITITPSPSQSTESTQTSLSPPSPSPSTHSVS 2005

QY 1999 GGATATGOGAVAG--LSKLSNDGOW--FKINGS 2031

Db 2006 SREGTTMSGATGDKMSFLSGTGTTFVSFRGSS 2041

RESULT 14

203099 mucin, submaxillary - pig

N:Alternate names: apomucin

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000

C:Accession: T03099 #A00009; A28528; B29789

R:Eckhardt, A.E.; Timpte, C.S.; DeLuca, A.W.; Hill, R.L.

J. Biol. Chem. 272, 33204-33210, 1997

A:Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain

A:Reference number: Z14839; MUID:98070526

A:Accession: T03099

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-13288 <ECK>

A:Cross-references: EMBL:AF005273; NID:g2581863; PIDN:AAAC62527.1; PID:g2581864

R:Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.

J. Biol. Chem. 266, 9678-9686, 1991

A:Title: Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in

A:Reference number: A40009; MUID:91236743

A:Accession: A40009

A:Molecule type: mRNA

A:Residues: 12139-12167, 'T', 12169-13288 <EC3>

A:Cross-references: GB:M61883; NID:g454837; PIDN:AAA30998.1; PID:g164374

R:Timpte, C.S.; Eckhardt, A.E.; Abernethy, J.L.; Hill, R.L.

J. Biol. Chem. 263, 1081-1088, 1988

A:Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical seque

A:Reference number: A28528; MUID:8808170

A:Accession: A28528

A:Molecule type: mRNA

A:Residues: 12139-12167, 'T', 12169-12641 <TIM>

A:Cross-references: GB:M21174; GB:J03512; NID:g164321; PIDN:AAA30990.1; PID:g5523260

A:Experimental source: submaxillary gland

R:Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Toumadje, A.; Johnson Jr., W.C.; Hill,

J. Biol. Chem. 262, 11339-11344, 1987

A:Title: Structural properties of porcine submaxillary gland apomucin.

A:Reference number: A92606; MUID:87280230

A:Accession: B29789

A:Molecule type: protein

A:Residues: 1572-1607 <EC2>

C:Superfamily: pig submaxillary mucin

C:Keywords: tandem repeat

Query Match 4.8%; Score 491.5; DB 2; Length 13288;

Best Local Similarity 20.0%; Pred. No. 2.2e-10;

Matches 464; Conservative 303; Mismatches 970; Indels 579; Gaps 96;

QY 8 IFNKATCTMAVAEYAKSHSTGGSCATCGVS--VCTLSFARIAALAVLVIGATLGS- 64

Db 1492 IYLOSTGISTBEE---SGSTGSLVSGKTSGLTILVTSRSSVAGSGTG-TVSGAS 1547

QY 65 -----AYAKQKDKHIAI-GEQNOPR-RSGTAKADGDRAITAGENANAOG- 107

Db 1548 GSTGSSSGSPGATRASIGQKTRRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGV 1607

QY 108 -----GQAIAGSSNKTVNGSSLD-----KIGTDATGQESIA-IGGDVKASGD 149

Db 1608 TETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSIRISVAGSSGAPAVSSG 1667

QY 150 ASTAIGSDLLHLLDQHGNPKRPHGTILNDLNGHAYLKBEIRSKDNDVKYRRTTAS-GHA 208

Db 1668 ASQAAGTSG-----AGP---GTTASSV---GVTETARPS-----VAGSGTGTGTVSGA 1708

QY 209 STAVGMSYACGHFSNAGFTRATKSNYSIAVLGAATAEGOS-----TIAIGSDATSSSLG 264

Db 1709 SGSTGSSSGPGATGASIQPETSIRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVG 1768

QY 265 AI-----ALGACTRAQLOG-----SIALGOGSVVQTQSDNNRSPAYTP 301

Db 1769 VTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSIRISVAGSSGAPAVSS 1828

QY 302 NTOALDPKFOATNNTKAGPLSIGSNSIKRKIIINVAGVKNKTDVAVVA--QLRAVYKWAKE 359

Db 1829 GAS-----QAAGTSGAGPGTTASS-----VGVTETARPSVAGSGTGTGTVSGASG 1872

QY 360 RRITFGDDNDSTDKIGLDNT--LTIKGGAETNALDNNIGVVKEADNSGLKVLAKTLN 417

Db 1873 STGSSSGSPGATGASIQPETSIRISVAGSSGAPAVSS--GASQAAGTSGAGPGTTASSV 1929

QY 418 NLTEVNTTTL--NATTTVKVGSSSTTAEILLSLST--FTQPNST-----GSQSTSKTV 466

Db 1930 GVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSIRISVAGSSGAPAVS 1989

QY 467 YGVNGVAFPTNNA---ETTAAIGTTRITRDKI-GFARDGDVD-----EKQAPYLD 511

Db 1990 SGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGAS 2049

QY 512 KKOLKVGSAVITIDNGIDAGNKKISNLAKGSSANDAVTIEQL---KAAKPTLNAGAGISV 568

Db 2050 IQPETSIRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSV-AGSGTTG 2108

QY 569 T-----PTEISVDKSGNVTAPTNYIGVKT----- 593

Db 2109 TVSGASGSGTSSSGSPGATGASIQPETSIRISVAGSSG--APAVSSGASQAAGTSGANGP 2165

QY 594 -TELSNDGTSD--KFSVKGSGTNNSLVTAEE-HLASYLENVEVRTADSALQSTVVKKEDDDD 649

Db 2166 GTTASSVGVVTETARPSVAGSGTGTVSGASGSGTSSSGSPGATGASIQPETSIRISVAGS 2225

QY 650 ANAIVTAKDITKNAG-----AVSILKLGKGLVATKKGDTVTFFGLSQDSGLTIG 700

Db 2226 SGAPAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGTGTGTVS-GASGSGSGSSG 2284

QY 701 KSTLNDGLTGVKDTNEQIQVGANGI-----KFTYNGSNPCT-----GIANTARTIR 747

Db 2285 SPGATGASIQPETSIRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTEA---R 2341

QY 748 DKIGFAGSDGAVDTNKNPYLDQDKLVGNVKNVTNTGINAGGKAITGLSPTLSIAQDSRN 807

Db 2342 PSVAGSGTGTGTVSGAS-----GSGTSSSGSPGATGASIQGP-----ETSRI 2382

QY 808 IELGNTIQDKDKSNAASINDILNTGFNLKNNNPIDFVSTYDIVDFAN---GNATTATV 863

Db 2383 SVAGSSGAPAVSSGASQAAGTSGAG-----PGTTASSVGVVTETARPSVAGSGTGTGTV 2434

QY 864 THDTANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTTKLNTSANGNTATNFNVNSDE 923

Db 2435 SGASGSGTGS-----SSGSPGATGASIQGP---ETSIRISVAGSSGAPAVSSGASQAAG 2483

QY 924 DALVNAKDIAENLNTLAKEIHTTKGTADTALOTFTVKKVDENNNDADNAITVQGNANN 983

Db 2484 TSGAGPGTTASSVGVVTETARPSVAGSGTGTGTVSGASGSGTSSSGSPGATGASIQGP---- 2539

QY 984 QVNTLTILKGENGLNITDKNGTVTFGINT---TSSGLKAKSTLNDGLGLKNP----- 1033

Db 2540 ETSIRISVAGSSGA-----PAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGT 2592

QY 1034 -----TGSEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTNGS--- 1080

Db 2593 TGTVSGASGSGTSSSGSPGATGASIQPETSIRISVAGSSGAPAVSSGASQAAGTSGAGPG 2652

QY 1081 -----LQSKPHPLSKDGIN---AGKKITNIQGEIAQNSHDVATGCKIYDLKTELE 1129

Db 2653 TTASSVGVVTETARPSVAGSGTGTGTVSGASGSGTSSSGS-----PGATGASIQPETSRI 2706

QY 1130 NKISSTAKTQNSLHEFSVADEQGNFTVSNPYSSYDTSKTS-----DVIITFAGENGIT- 1183

Db 2707 SVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVVTETARPSVAGSGTGTGTVSGASGSGT 2766

Db	725	TVVNATATDASGNTSNGSS-VTVDSVAPATPVINPSNGTITLSGTAEPGSSVTLTDGNGNP	793
Qy	750	IGPAGSDG----AVDTINKPYLDDKLQVGNKVIINTGINAGGKAIT---GLSPTLPSTI--	800
Db	784	ICQVTTADGSSNWSFTSPSTPLADG---TVVNATATDPAGNTSGOGSTTVVDGVAPTPTVNL	840
Qy	801	--ADQSRNIELGNTIQDKKSNAASINDILNTGFKLNKNNNPIDFVS-----TV---	848
Db	841	SNGSSLSGTAEPGSTV-----ILTDG-----NGNPAEYTAGDSGNWTPTPS	892
Qy	849	-----DIV---DFPANGNATTATVTHDTANKTSKVYDVNVNDDTTHILHTGDDKNKLG	897
Db	883	TPITANGTVVNVVAQDAAGNSPGASVTVDSSQAAPVVPVPSN--GTT--LSGTAEP--G	935
Qy	898	VKTTKLNKTSANGNATNFWNNSD-----EDALVNAKDIAENLNTLAKELHHT	946
Db	936	ATVTVT---LTDGNGNPICQVTAGDGSNWSFTPTPLANGTVVVNATASDPTGNTSAPASTTV	992
Qy	947	KGTADTALQTFVKVYDENNNAD-----DANAITVGQKNANNQVNTLTLXGELNLIKTD	1001
Db	993	DSVAPAA-----PVVNPNSGAELISCTAEPGATVTLTDGSCNPICQVTDGSGNWSFTPS	1046
Qy	1002	---KNGTVTFGINWTSGLKAGKSTINDGSLSTKNPTGSEQIQVGADGVKFAKVVNNNGSVG	1058
Db	1047	TPLADGTV---VNATATDPAG---NTGG-----QGSTTVDAIAPATPTVNLN---G	1089
Qy	1059	AGTDGTRITRDRIGTGTNGSLDKSKPHLSKDGINAGGKXITNIOGSEIAONSHDAVTG	1118
Db	1090	SSLSGTA-----EPGST-----VILTDGNGNPIDAEVTDAGSGNWTPTPTPIAN	1133
Qy	1119	GKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNNFTVSNPSYSDTSKTSD-----V	1173
Db	1134	GTVVNVVAQDASGNSSPPATV-----TVDSSAPPAPVILNPNSGVVISGTAEGAGATVT	1185
Qy	1174	ITAGENGII-----TTKVNKGVVRVGIDQTKGLITPK--LTVGNNGNGK	1215
Db	1186	LTDAGNPIQOVTDAGSGNWSFTPTPLANGTVIVATATDPTGNTGPOAATTVDAVAPPA	1245
Qy	1216	IVTDSONGQNTITGLSNTLIANTYNDKGSVRTTEOGNIIFKDBDKTRAASIVDVLSAGPNLQ	1275
Db	1246	PVIDPSNG-TTISGTAEGAKV-----ILTDGNGNPI--GETTADSGNWSFTPGTPL--	1295
Qy	1276	NGEAVDFVSTYDTVNFDAGNATTAKVYYDDTSKTSKVYVDNVVD-----DTTIEVK	1327
Db	1296	ANGTVVNAQAQDPAGNTGPQGSTTV---DAVAPTPVNPNSGNLLNGTAEPGSTVTLT	1351
Qy	1328	DKLGVKTTTLTGTCGANKFALSNOATGDALVK--ASDIVAHLN---TSLGSDIQTAQ	1380
Db	1352	DGN-GNPICGOTTADSGNWSFTPGSQLPNGTVVNVYATSDAAGNTSLPATITTVDSLSIP	1410
Qy	1381	GASOANNS--AGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKKEVKDKLVAAQAQTPDGT	1438
Db	1411	QVDPNSGVSISGTPADAGNTIITLDGNGNPICQVNTADGS-----GNWSFTPGIPLPDGT	1463
Qy	1439	LAQMWKYSVINKEOVDANKKOGINEDNAFV---KGLEKAASDNKTKNAATVYCDLNAVA	1495
Db	1464	VNVVARSNPVDSAPAVITVDGVAPAPVDPDSNGTSEISGTAEGATVILTDGGGNGPIC	1523
Qy	1496	QTPLTTFAGD--TGTTAKKLGTELITIKGGQTDTKNKLTDNNIGVVAQTDGFTV--KLAKDLTNL	1553
Db	1524	QATADSGNWTETPTSPLANGTVI-----NAVAQDPAGNTSGPASVTVDATAAPPAPVI	1576
Qy	1554	NSVNAVGTCKIDDKGVSEFVDSGQAQKANTPVLSANGLDLGGKVISNV---GKG-----TKDT	1606
Db	1577	NPSNG-----VVISGTAEGATVILTDG---NGNPIQOVTDAGSGNWSFTPGT	1621
Qy	1607	DAANVOQLNEVRNLLGLGNAGNDNADGNQVNTADTKK-----DPNSGSSSNRTYKIKAGTV	1661
Db	1622	PLAN-----GSVINLAQDAAGN-NSPSTSATVDSLAPAAPVIDPSNGSVIAGTAEAGATV	1676
Qy	1662	LGKGKGNNDTEKLTGGTQGVGVKDGDN-----ANGDLSNVWVKYTKQDGSKKALLATY	1712
Db	1677	ILTDGNGN---PIG--QVTDAGSGNWSFTPGTPLSNGTGVVNA-----VAQ	1716

Search completed: September 13, 2001, 12:39:59
Job time: 32665 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:39:30 ; Search time 47.77 Seconds
(without alignments)
1467.887 Million cell updates/sec

Title: US-09-361-619-7
Perfect score: 10303
Sequence: 1 MNHIYKVFNKATGTFMAVA.....NGSADTQGHVGAAGAGPHF 2047

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	627	6.1	2249	1 190K_RICRI	P15921 rickettsia
2	506	4.9	2003	1 YDBA_ECOLI	P33666 escherichia
3	471	4.6	1577	1 HLYA_PROMI	P16466 proteus mir
4	469.5	4.6	1654	1 OMPB_RICRI	Q53047 r outer mem
5	457.5	4.4	1643	1 OMPB_RICRY	P96989 r outer mem
6	456	4.4	1643	1 OMPB_RICPR	Q53020 r outer mem
7	451.5	4.4	1902	1 P2P_LACLA	P15293 lactococcus
8	448	4.3	3591	1 FHAB_BORPE	P12255 bordetella
9	447.5	4.3	1902	1 P2P_LACPA	Q02470 lactobacill
10	443.5	4.3	1656	1 OMPB_RICJA	O06653 r outer mem
11	441.5	4.3	1902	1 P1P_LACLC	P16271 lactococcus
12	430.5	4.2	1902	1 P3P_LACLC	P15292 lactococcus
13	417	4.0	1608	1 HLYA_SERMA	P15320 serratia ma
14	375	3.6	1325	1 YDEK_ECOLI	P32051 escherichia
15	370.5	3.6	2334	1 WAPA_BACSU	Q07833 bacillus su
16	366.5	3.6	1569	1 YPJA_ECOLI	P52113 escherichia
17	360.5	3.5	1848	1 CBPA_CLOCL	P38058 clostridium
18	355.5	3.5	1300	1 120K_RICRI	P14914 rickettsia
19	351	3.4	1567	1 ICEN_XANCT	P18127 xanthomonas
20	339.5	3.3	1286	1 AIDA_ECOLI	Q03155 escherichia
21	329.5	3.2	1176	1 SLAP_BACSH	P38537 bacillus sp
22	323.5	3.1	1849	1 IGAA_ENTFA	P45386 haemophilus
23	312.5	3.0	1296	1 ASAL_ENTFA	P17953 enterococcu
24	310.5	3.0	1025	1 SLAP_CAUCR	P35828 caulobacter
25	309.5	3.0	918	1 YMJB_CAEEL	P34487 caenorhabdi
26	305.5	3.0	1702	1 IGA2_HAEIN	P45384 haemophilus
27	304.5	3.0	1592	1 GTF2_STRDO	P27470 streptococc
28	301.5	2.9	1694	1 IGAA_HAEIN	P44969 haemophilus
29	301.5	2.9	2628	1 HAGA_PORGI	Q51845 porphyromon
30	299	2.9	1829	1 FRPC_NEIMB	Q9JYV5 neisseria m
31	298	2.9	1419	1 ALA1_CANAL	O13368 candida alb
32	294	2.9	1228	1 SLAP_BACST	P35825 bacillus st
33	292	2.8	1183	1 CNA_STAAU	Q33654 staphylococ

34	291.5	2.8	1829	1 FRPC_NEIMC	P55127 neisseria m
35	291	2.8	1039	1 AG43_ECOLI	P39180 escherichia
36	286.5	2.8	1243	1 VG37_BPK3	Q38394 bacterioph
37	286.5	2.8	1541	1 IGAL_HAEIN	P42782 haemophilus
38	285.5	2.8	933	1 SLAP_CAMEE	P35827 campylobact
39	285.5	2.8	1322	1 ICEA_PANAN	P20469 pantoea ana
40	285	2.8	1637	1 MRSP_STAAU	P80544 staphylococ
41	284.5	2.8	827	1 CSG_HALVO	P35062 halobacteri
42	283.5	2.8	1258	1 ICEN_ERWHE	P16239 erwinia her
43	283	2.7	1861	1 APU_THETU	P38536 t amyllopull
44	280	2.7	1462	1 GTFD_STRMU	P49331 streptococc
45	278	2.7	1140	1 YM96_YEAST	Q04893 saccharomyc

ALIGNMENTS

```
RESULT 1
190K_RICRI
ID 190K_RICRI STANDARD; PRT; 2249 AA.
AC P15921;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
  repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990).
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -!- PTM: GLYCOSYLATED (POSSIBLE).
CC -!- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
  AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL; M31227; AAA26380.1; -
CC PIR; A41477; A41477.
CC Antigen; Repeat; Signal; S-layer; Glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 2249
FT DOMAIN 212 1180
FT REPEAT 212 286
FT REPEAT 287 358
FT REPEAT 359 430
FT REPEAT 431 505
FT REPEAT 506 577
FT REPEAT 578 652
FT REPEAT 653 724
FT REPEAT 725 799
FT REPEAT 800 874
FT REPEAT 875 949
FT REPEAT 950 1021
FT REPEAT 1022 1093
FT REPEAT 1094 1165
FT REPEAT 1166 1180
FT REPEAT 1180 1180
SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;
```


Query Match		6.1%;	Score 627;	DB 1;	Length 2249;
Best Local Similarity		22.2%;	Pred. No. 1.6e-15;		
Matches 477;		Conservative 233;	Mismatches 754;	Indels 684;	Gaps 108;
QY	97	IAIGENA-----NAOQGOAIAIGSSNKTGVSSLDKIGTDATEQCE---SIAIGGDVKA	146		
DB	44	IATNNAAFSNVGNWNNWNEITAAGVAN-----GTPAGGPQNWAFTYGGDYTV	92		
QY	147	SGDASTAIGSDDLHLHDQHGNPKPG-----TLINDLI-NGHAVLKEIRSSKNDVDKY	199		
DB	93	TADAADRI-----IKAINVAGTT--PVGLNITQNTVVGSIITKGNLLPVTLAGKSLTLNG	146		
QY	200	RETTASGH-----ASTAVGAMSVAOGHESNAGTRATKASVSLAVGLAATAEGOSTIAI	254		
DB	147	NNAAAHGCFADAPADNYTGLGNIALGGANAALIIQSAAPS-----KITLAGNIDGGIITV	202		
QY	255	GSDA-----TSSSIIAGALGAGRAQLOQSI-----ALGQGSVVTQSDNNRPPAYT	300		
DB	203	KTDAANGTIGNTINALATVNVGAGT-ATLGGAVIKATTTKLTNAASVLTLTNANA-----	256		
QY	301	PNTQALDPKFOATNNTKAGPLSIG-----SNSIKRKLINVGAGVKNKTDAVNAQL	350		
DB	257	VLTAID-----NTTGGDNVGVNLINGALSQVTDGIGNTSLATISVGAG-----TATL	305		
QY	351	EAVVWAKERRITFOGDDNSTDVKIGLDNTLTIKGGAETNALTDNNIGVYKADNSGLKV	410		
DB	306	GGAVIKATTTKLT-----DAASAVK--FTNPVVVTGAIDNTGNANN--GIVTFTGNSTVTG	357		
QY	411	KLAKTLNLTNEVNTLTNATTVKVGSS-----SSTTAELL--SDSLFTPTOPNTGS	459		
DB	358	NVGNNT-----NALATVNVGAGLQVQGVVVKANTINLTNANASAVTFTNPVVVT	405		
QY	460	QSTSKTVGVNG-VKFTNNAAETTAICTRITRDKIGFARGDGDVDEKQAPYDKKOLKYG	518		
DB	406	CAIDNTGNANGIVTFTGNTSTVTDIGNTNAL-----ATVNVG	443		
QY	519	SVAITIDNG-IDAGNKKISLAK-----GSSANDAVTTEQLKAAKPTLN	561		
DB	444	AGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLINGALSQVT	503		
QY	562	AGAGISVTPTEISVDKSGNVTAFTYNGVKTTELNSDGTSDKFS-----VKGS-----GT	612		
DB	504	GNIGNTSLATISVGA--GTATLGGAVIKATTTKLTDAASAVKFTNPVVVTGAIDNTGNA	561		
QY	613	NNSLVTAHLASLYNEVNTADSALOSFTVKEEDDDANAITVAKOTTKNAGAVSILKIK	672		
DB	562	NGIVTFTGNTSTVTDIGNT--NSLATISVGAGTATLGGAV-IKATTTKLTNAASVLTLT	618		
QY	673	GKNGLTVAATKKDGT-----VTFGLSDSG-----LTIGKSTLNDGLTVKD	713		
DB	619	NANAVLTGAIDNTTGGDNVGVNLINGALSQVTDGIGNTNSLATISVGAGTATLGGAVIKA	678		
QY	714	TNEQIQVGANGIKETN---VNGSNPGTGIANBARITRDKIGFAGSDGADVNTKPYLDQDK	770		
DB	679	TTTKITNAVSAVKFTNPVVVTGAIDSTGNANGIVT-----FTGNSTVTG-----	723		
QY	771	LOVGNVKTNTGICNAGGKAITGLSPTLPSIADOSSRIELGNTI---QDKDKSNAASIND	827		
DB	724	-DIGNTNALAT--VNVGAGTAT-----LGGAVIKATTTKLTNAASVLT	763		
QY	828	ILNTGFENLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYDYVNVDDTTHL	887		
DB	764	LTNANAVL---TGAIDNTTGGDNVGVNLINGALSQVTDGIGNTNSLATISVGAG-----	814		
QY	888	TCTDDNKKLGKVTTLKLNK-----TSANGN---TATNFVNNSDEDALVNAK-----D	931		
DB	815	TATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDNVGVNLINGALSQVTDG	874		
QY	932	IAENLTLAK-----EHTKTGDTADALQOTFTVKKVDENNADANAT--	974		
DB	875	IG-NTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTLT-----NANAVLTGAIDN	927		
QY	975	-----TVGQKNANNQVNTLTLLKGENGLNIKTDKNKNTVTFGINTTSGLKAGKSTLNDGGLS	1029		

DB	928	TTGTDNVGVNLINGALSQVT--GDIG---NTNSLATISVG-----AGTATL--GGAV	972		
QY	1030	IKNPTSEQIQVQCADGVKPAKVNNGVVGAGIDGTTTRITRDEIGFTGNT---GSLDKSRP	1086		
DB	973	IKATT--TKLTDAAAVKF--TNPVVVTGAIDNTGNANGIVTFTGNTSTVTCNVGTNA	1027		
QY	1087	HLSKDGINAGGKITTNIQSGEIAQNSHDVAVTGKIIDYDLKTELENKISSAKTAQNSLHEF	1146		
DB	1028	-LATVNVGAG--LLQVOGVVKANTINLRD-----NASAVTFTNPVVVT	1068		
QY	1147	SVADEQGNNTVSNPYSSYDTSKTSVDITPAGENGIT--TKVNGVYVVRGVIDQ	1197		
DB	1069	GAIDNTGN-----ANNGIVTFTGNTSTVTCNVGTNATVNVG---AGLLQ	1111		
QY	1198	TKG-----LTPPKLTVG-----NNNGKIVIDSQNGQNTITG--LSNT	1233		
DB	1112	VOGVVKANTINLTNANASAVTFTNPVVVTGAIDNTGNANGIV--TFTGNTSTVTDIGNT	1169		
QY	1234	LANYNDKGSVTRTEQGNIIKDEDKTRAASIVD-----VLSAGFNLOGNGEAVDF-----	1283		
DB	1170	NALATVNVGAGITLQAGGSL-----AANNIDFGARSTLEFNGPLDGGGKAIPYYPFGA	1222		
QY	1284	-----VSTYDITVNFADGNATTAKVYDTSKTSKVYVDVNVDD	1321		
DB	1223	IANGNNALNVNTKLLTASHLTIGTVAEINIGAGNLTIDASVGD-----VTILNAQN	1275		
QY	1322	TTIEVKDKKLVKVTTLTSTGTGANKFALSNOATGDALYKASDIVAHLNLTLSGDIQT---	1378		
DB	1276	INFRARDSVLVLSNL---TGCVGNINLL-----AADLVAP---GADEGTVPF	1316		
QY	1379	---AKGASQANNSAGYV---DADGNC---VIYDS---TDNKYVOAKNDGTVDKTEVA	1424		
DB	1317	NGGVNGLNVGNSVAGTARNIGGGGKENTLILYNAVITTDVNLEGIONVLINKNADET	1376		
QY	1425	K-----DKLVAQAOTPDGTL--AQMNVKSVINKEOVNDANKKQGINE	1464		
DB	1377	SSTAFNAGAIQINDATYITIDANNGLNLPAGNIQFAHADALVLQNSSGNDORTITLGANI	1436		
QY	1465	DNAFVKGLEKASDNKTKNAAVTVGDNAV-AQTPLTFAGDTGTT---AKKLGFTLTITKG	1520		
DB	1437	D-----PDNDEGIVI---LNSVTAGKLLTAG--GKTFGGAHKL-QTILFKG	1478		
QY	1521	-GQTDITNKLTNNIGVAGTGDGTVYKLAKDLTLNLSNVNAGGTKIDDKGVSVFSDSQAKA	1579		
DB	1479	AGDCST-----AGTFTNTTNIULDIT-----GQ---	1501		
QY	1580	NTPVLISANGLDLGGKVISNVGKTKTDAAVQOQLNEVRNLLGLGNAG---NDNAD---	1632		
DB	1502	-----LELG-----ATTANVVLFDNAVOLTQTGNIIGGLDFDNAGNMV	1539		
QY	1633	--GNOVNIADIKKDPNSGSSSNRTVIKAGTVILGGKGNNDTEKLA---TGGIOGVVDKQGN	1687		
DB	1540	TLNNVNVVAGVQ--NTGCTNGTLLI---VLGASNLNRVNGIAMLVKVGAGNVTTAKGK	1593		
QY	1688	AN-GDLSUNVWVKQDKGKALLATYNAAGQNTLYLTNNPAAEIDRINEOGIRFFHVDGN	1746		
DB	1594	VKICEIGQGTGNT-----LTLPAHFNLTG---SINKTGGQALKLNPNMGSS	1636		
QY	1747	QEPVVOGRNGCIDSSAGSKHSAIGFOAKADGEAAVAGRQTOAGNQSIAIGDNAQT---	1803		
DB	1637	VSGV-----GTAANSVGDITTAGATSFASSVNAKGTATL	1671		
QY	1804	-GQOZIA---IGTGNVNVAGKHSIGAIDPPSVKADNSYSVGNNNQFTDQTDFVGVGNNI	1859		
DB	1672	GGTTSFANTFTNTGAVTLAK-----GSITSFAKNVTATSF--VANSA	1711		
QY	1860	TVTESNVALGNSATLSAGTHAGTQAKKSDGTAGTTTGTAG-----TGTVKGFAGGTAVG	1914		
DB	1712	TINFNSLAFNSN--ITGG-----GTTILGANQVYTTGT--GSFTDITLTLN	1754		
QY	1915	AVSVGA--SGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQSTANATN	1960		

Db 1755 TTFDGAAGSGNLIKSGTLDLSCVSTLAL-----VVTATN 1791

RESULT 2

YD8A_ECOLI STANDARD: PRT: 2003 AA.
AC P33666: P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
GN YD8A OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RA "The complete genome sequence of Escherichia coli K-12";
RT Science 277:1453-1474(1997).
RL (2)
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.,
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RT DNA Res. 3:363-377(1996).
RL (3)
RN SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RA "Multiple IS insertion sequences near the replication terminus in
Escherichia coli K-12";
RL Biochimie 73:1361-1374(1991).
CC -1- SIMILARITY: TO S.TYPHIMURUM ORF NEAR CYSG (AC P25928).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000237; AAC74483.1; ALT_SEQ.
DR EMBL: AE000237; AAC74487.1; ALT_SEQ.
DR EMBL: D90778; BAA15009.1; ALT_SEQ.
DR EMBL: D90778; BAA18880.1; ALT_SEQ.
DR EMBL: D90779; BAA18881.1; ALT_SEQ.
DR EMBL: X62680; -: NOT_ANNOTATED_CDS.
DR EcoGene: EG11307; yd8A.
DR Hypothetical protein.
KW CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.9%; Score 506; DB 1; Length 2003;
Best Local Similarity 21.0%; Pred. No. 3e-11;
Matches 423; Conservative 264; Mismatches 775; Indels 554; Gaps 96;

Qy	307	DPKFOATNTRAGPLSGNSIKRKLIINVA--GVNKTDAVNVAQLEAVVKA-----	357
Db	159	DSVFTYTENAD-GTISLQDSNGRKATINLWQIDEANNTVALEGVGADGATKQYHNHNGEL	217
Qy	358	-----KERRITFOGDD-----NSTDVKIGLDNTLTIKGGAETNALTDNNIGV--	399
Db	218	VITGDATVNNGKTTVDGKDGSTGTEINGNCKKVIQDGLDVSFGGHHGIDITGDSATVDN	277
Qy	400	-----VKADNSGLKVLAKTLNLTNTVNTLTATTTVKVGSSTSTAEALLSDSLTFTQ	454
Db	278	KGTWFTVDPESMGVQIDGDKAIVN-NEGESTITNGTGTQI-NGDDATANNNGKTTVDGK	335
Qy	455	PNTGSO---STSKTVY-----GVNGVKPTNNAETAAIGTTRIT-RDKIGFARDGDV	502
Db	336	DSTGTEINGNCKVIQDGLDVSFGGHHGIDITGDSATVDNKGTTMTVDPESIGIQVDGD-	394
Qy	503	DEKQAPYLDKQKLVGSAVITID-----NGIDA---GNKKISNLAKGS-----SANDAV	548
Db	395	---QAVVNE-----GESAITNGTGTQINGDDATANNNGKTTVDGKDGSTGTEIAGNCK	446
Qy	549	TIEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTPATYINIGVKVTELSNDSGTSKFSVK	608
Db	447	VIQD--GDLDVSGGHHGIDITGDSATVDNKGTTMTVDPESIGIQI-----DC--DQAI	497
Qy	609	SGTNNLSLVTAHLASYLNEVNRATDSALQSFVKEEDDDANAITVAKDTTKNAGAVSI	668
Db	498	NEG--ESTITNGTGTQINGDDATANNNGKTTVDGKDGSTG-----TKIAGNIGI	544
Qy	669	LKLKGKNGLTATKKDGTVTGSLQSDSGITIGKSTLNNDG-LTVKDTNE-QIOVGANGTK	726
Db	545	VNLDC--SLTIVTGAHGVENIG---DNG-----TVNKGDIVSDTSGSIGVLINGEGAT	593
Qy	777	FTN---VNGSNPCTGIANTRITRKIGFAGSDGAVDINKPYLDQDKLOVGNKVTNTGI	783
Db	594	VSNTGDVNVSEATGFSITT---NSGKVSLAGS-----MQVGDFS-TGVDL	635
Qy	784	NAGGKAITGLSPTPSI-----ADSSRNIELGNTIQQDKS--NAASINDILNT	831
Db	636	NGNNNSVTLAAKDLKVGQKATGINVSGDANTVNI-TGNVLVDKDKTADNAAEYFDFSV	694
Qy	832	GFNLKNNNNPIDFSTYDIVDFANGNATTATVTHDTANKTSKVYDVNVDDTTIHLTGTD	891
Db	695	GINYVGDNNVTLDGKLTWVSDSEVTSGSNLFDGSAEKTSGLV--VIGDGNVTNMG--	750
Qy	892	DNKKGKVTTLKNTSANGNTATNFVNSSDEEDALVNAKDIAENLN---TLAKEIHTTKG	948
Db	751	-----GLELIGEKNALADGSQVTSRTGYSTSVIVSGESSVYVNLGDDTTISGEF--	803
Qy	949	TA-----DTAL-----QTFVKKVDENNADANAITVGQKNANNQVNTLTLAGEN-G	995
Db	804	FAGVTRVDQKALLETSATLTMQDIDSFHH-----GTRTVEIQNLG	846
Qy	996	LNKTKDKNGVTFTGINTTSGLKAGKSTLNDGSLSKNPTGSEQIQVGADGVFAKVNNG	1055
Db	847	FAFVTGEN-----ITGINSGTISLQNG---KDPAPSPVILLATNG---GSATNAG	891
Qy	1056	VVGAGIDGTTTRTRDEIGF-----TGTNSLDSKSPHLSKDGINA-GGKKITNQSGEIAQ	1110
Db	892	TI-----TGKVTQHSFVNKYSTGTSTNSFIFNNDVSSITGLVAOSNSTIINTDSGIDL	945
Qy	1111	NSHDAVTGKGLYDLKTELENKISSAKTAQNSLHEFSVADEOGNNTVSNPSSYDTSKT	1170
Db	946	YGRSGVGLATAD-----STAENQGGKITLDSMWVDANDTTAM	982
Qy	1171	SDVITFAGENGITTKVNGKV-VRVGIDQTKGLTTPKLTGVGNNGKGVIDSQNGQNTIIG	1229
Db	983	RDIAS-----NSAIDFGTGVGVGTDVSG-AGKNATAINQLGGVITIIYNAGAGMAAYG	1034
Qy	1230	LSNTLANTVNDKGSVRTTEQGNIIKDEKTRAASIVDLVSAGFNLOGNGEAVDFVSTYDT	1289

3

Db 350 QGADLTIDGKOLQKQETDIDNRHFPYSWKYDVTKKEQIQIGSQIDAKNNATLTATKGDV 409
QY 911 -----NTATFNVSNDSE-----DALVNAKDIAENLNLAKIEHTTK-----GTADTALQTF 957
Db 410 TLDAAKINAGNLAIAINANKDIHINGLVEKESRENGN---KRHTSRLESQSNSSHOTE 466
QY 958 TVRKVD-----ENNADANAIV-GOKNANNQVNTLTKGENGUNI-KTKNGTIVTFCIN 1011
Db 467 TLKASELTAGKDLGLDAQSGITAGAKLANENVLVNAKDNIINLVQKTNNDKTVTDNHV 526
QY 1012 TTSGLAKGKSTLNDGGLSIKNPQSGEIQVQAGDS-VKFAKVNNGVVGAGIDGT-----TR 1066
Db 527 MWGIGGQNKNNNOOVSHAT-----QUTADQQLLLAADNNVNITGSOVKNGQGAFAVK 581
QY 1067 ITRDEIGF-----TGNGSLDKSKPH-----LSKD 1091
Db 582 TTQGDVVIDNALSETISKIDERTGTAFNITKSSHKNETKQTSQSGELISDAQLTVVSGN 641
QY 1092 GINAGKKITNIOS-----GEIAQNSHDVATGKGIYDLKTELENKISSAKTAQNSLHE 1145
Db 642 DVNVIGSLIKSADKLGTHSLGDIINVAQOVT--KIDDEKTSL--AITGHAK-----689
QY 1146 FSVADEQ---GNFT-VSNPYSYDTSKTSVITFAGENGITTKVNGKVVVRVGIDQTKGL 1201
Db 690 -EVEDKQYSGAFHHTTNKNTSTETEANS--TISGAN-VDLQANKDVTFAGSDL-----741
QY 1202 TTPKLTGVNNN--GKGIVI-----DSQNGQNTITG-----LSN 1232
Db 742 --KTTAGNASITCDNVAFYSTENKQTDNTDTISGFSYTGVDKVGSKADFOYDKQH 798
QY 1233 TLANVNDKGSVRTTEOGNLIKEDKTRAASIVDVLSAGFN-----LQNGEAVDPVST 1286
Db 799 TQTEVTKNRGS-QTEVAGDLTITANK-----DLLHEGASHIHVEGRYQESGNIQHLAV 850
QY 1287 YDTVPNADGNATTAKVTYDFTSTSK-----VVVDVNVDDTTIEVKDKKL--GVKTTTLT 1339
Db 851 NDS-----ETSKTDSLNVGIDVGNLDYSGVTKPVKKAIEDGVNTT---891
QY 1340 STGTGANKFALSNOATGDALVKASDIYVAHLNTLS-----GDITAKGASQ 1384
Db 892 ---KPGNNTDLTKKVT-----ARDAIANLANLSNLETPNPGVEVGIKGGSQSOQSIDSQ 942
QY 1385 A---NSAGVVDAGNKVIVYDSTDNKYQAKNDGTVDKTEKAKDKLVAQAQPDGTL--1439
Db 943 AVTSINAGRIDISNNKLHD--QGTHYQSTQEGI-----SLTANTHTSEATLDK 990
QY 1440 -----AQMNVKSVINKQVNDANKKOGINEDNAFVKGLEKAASDNKTKNAAVTV 1488
Db 991 HQTTFHETKGGQIGVSTKTKGSD--ITVAIKGEGOTTDNALMETKAKGQTSNGDISINV 1049
QY 1489 GD-----LNVAQOTPLTFAGDT-----GTTAKKLGETILTIK---1519
Db 1050 GENAHYEGAQDAQKQKTVINAGGDLTLAQATDTHSESQSQNVNCSANLKVTGTPESKDYG 1109
QY 1520 ----GGQTDNKK-LTDNNIGVAGTGDFTVKLAKDLT-----NLNSVNAAGTK 1562
Db 1110 GGFNAGTTHHSKEQTAKVQITGTSQGIENLGNHNLTLQGHLSSEQDIALNATN----K 1165
QY 1563 IDKGVSVFVSSGOAKANTPVLSANGLDIGKVIKSNVKGCTKDTDAANVOOLNEVRNLG 1622
Db 1166 VDLQAS--SEHTEKGN-----NLSSGGVOAGFGKAKMTD--DASSVNGLSAQAPAG 1212
QY 1623 -----LGNAGNDNADGN-----QVNTIADIKKPNSSG-----SSSNRTVIKA 1658
Db 1213 KODEKSVSREGGTINNSGLNTINGNSVHLOGAQNKSQDTQITQSQGDIEITSQAQSTYDN 1272
QY 1659 --GTVLG--GKGNNDT-----EKLAT-----GGIQGVGD-----KGNANGDLSNVWV 1697
Db 1273 NWGTDIGFNGKKNTNTPKEVTEKPAATSIHIGGKLLVNVVEDQOKTSHQNTLETGTITI 1332
QY 1698 KTOKD-----GSKKALLATYNAACQTNLTNNPAEATDRINEQIR--FFHVNDGNGEPV 1750

Db 1333 NSNKDLTSLGANVTADSVTGNVGGSLNIASOKES---DRHVTVGNNVGYNTNDPKSSQV 1389
QY 1751 VQERNGIDSSASG-----KHSVAIGFOAKADGEAAVAIGROTAQAGNSIA-----IGDN 1799
Db 1390 NK-----TAKAGGSLLKTIKDTIDSGIKSSTD-----AISDKYNSLSSTIADTKTGSDE 1439
QY 1800 AQATGDOSIA-IGTG--NVVAGKHSAGIADPSTVKADNSYSVGNNOFTDATQYDVFVG 1856
Db 1440 TKAKIDGFGKVGNGIKNIIVTG-----AEGHTANADIKVTHVDNDVATKTT-----1485
QY 1857 NNITVTSNVALGNSAISAGHTAGTQAKSDCTAGTTTAGATGTVKGFAGOTAVGAV 1916
Db 1486 ---SLTSNNDLSLVN-----GSTKLTGA--EIVSQQGVQVDLGG 1520
QY 1917 SVGASGAERRIQNVAAAGEVSATSDAVNGSOLYKATQSIAN-----ATNE 1961
Db 1521 SV-----KLENEGHHYEAGADLDLSSVVVDLAKQLVGGDISFKSPVKINE 1566

RESULT 4

OMP_RICRI
ID OMPB_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE [SURFACE PROTEIN ANTIGEN] (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (romp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X16353; CAA34403.1; -
DR Antigen; S-layer; Cell wall.
KW CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
SQ

	Query Match	4.1%	Score 469.5;	DB 1;	Length 1654;
	Best Local Similarity	22.1%;	Pred. No. 4.8e-10;		
	Matches 386;	Conservative 174;	Mismatches 561;	Indels 629;	Gaps 90;
y	650	ANAITVAKDTRKNAGVASILKKGKNGLTIVATKKGDTVFGLSQDSGLTIGKSTLNNDGL	709		
b	19	ASTATVASFAGSAGAAIQONRTTG--AAITVDGA--GFDQTAAPA-----NVGV	66		
y	710	TVKDTNEQIQVGA-NGIKTINVSGNPGHGIANTAR-----ITRD-KIGFAGSDGAVDTN	762		
b	67	AL--NAVITANANGINGENTPAGSENGL-LLNTANNLAVTVSEDTTLGFI-----TN	115		
y	763	KPYLDQDKLVGVNKKITNTGINAGGK-ATIGLSPTLPSTADOSSRNIELGNTIODKRSN	821		
b	116	-----VVHNAHSFNLTLNAGKTLTITGCGVTNAQAA--ATKNAQ--NVVGFNNGA	162		
y	822	AASINDILNTG-----FNLKNNNPIDFVSTYDIDVFDEAGNAGTATVTH-----D	866		
b	163	AIDNNDLKGVRIDFGAPASTLVFNLANPTQAKPILIGDNVANGVNGTLNVTNGFIO	222		
y	867	TANKTSKVYDVNVDD-----TTIHULTGHDDNKKLGCVKTKTLN	904		
b	223	VSNKSEATVRAINADQGGIFINTDANNANTLQAGGTINTFGDTGGRL-----VLLS	278		
y	905	KTSANGTATNFNVNSDEDALVNAKDIAENLNTLAKEIHTTKGTADTALQTFTVKKVDE	964		
b	279	KHAA-----ATNFNTGSLGG---NLKGVIE-FNTVAVDGOLTA-----313			
y	965	NNNADDAITVVGOKNANNQVNTLTLKGENGLNIKDKTKGTVTFGINTVSGLKAGKSTLN	1024		
b	314	--NAGAANAV-IGTNNGAGRAAGVVSVDNG-----KVATID-----GVQYAK	353		
y	1025	DGGLSINKPTGSEQIOVGADGVKFAKVNNGVGVGADIGDTTRITROEIGFTNGSLDKS	1084		
b	354	DMVITQSANATG-----QVNFRIHVDVGADGIT-----380			
y	1085	KPHLSKDGINAGGKKIINIOGSETAQNASHDVAVTGGKGIYDLKTELENKISSTAKTAQNSLH	1144		
b	381	-----AFKTAASKVT-----ITQDSNFGNTDFGNLAQIKVPNAITLT-----418			
y	1145	EFSVADEQGNFT--VSNPYSSYDTSKTSVDVITF-----AGENGITTKVN	1187		
b	419	-----GNFTGDASN-----GNTAGVITTFDANGTLESADANVAVTNNTAIEA	463		
y	1188	KGVVRVGIDQYKGLTPTPKLTVGNNN-----CKGIVIDSQNGQNTITGLSNTLANVTNDKG	1242		
b	464	SG--AGVVQLSGTHAAELRLGNAGSTFKLADGTVINGKVNQTVALVGGAAGAAGTILDGS	520		
y	1243	SVRTTTEQGN-----IHKDEDKTRAASIVDVLISAG---FNLOGNGEAVDFSVSTYD	1288		
b	521	AITITGIGNAGCAAAALQRTITLANDAKKTLTGGANILGAGGCTIDLOANGCTIKLTSTQN	580		
y	1289	--TVNFDAGNATTAKVTVDDTTSKTSKVYDVNVDDTTIEVKDKKLGKVTTLTSTGTGGA	1346		
b	581	NTVVPFDLIAIATDQGVVDASLTL-----NAQTLTINGKIGTIGANNKTLGOFNTGSS	633		
y	1347	KEALS-NQATGDALVKASDIVAHLNLTSCDIQTAKGA---SQANNSAGYVDADGNKVIYD	1402		
b	634	KTVLSNGVAINELVIGND-----GAVQFAHDTYLITRTTNAAGO-----GKIIFN	679		
y	1403	STDNKKYQAKNDGTVDKTEKAVAKKLVAQAQPTPDGTLQAMNVKSVINKEQVNDANKKGI	1462		
b	680	-----PVVNNGT-----LAAGTNLGSATNP---LAEINFGS-----KGV	711		
y	1463	NEDNAP--VKGLEKAAASDNKTKNAAV-----TVGDLNVAQATP-----LTFAGDTGT	1507		
b	712	NVDTVLNVGEGVNLATNTTTDANVGSFVNAGGTINVSQTVGSGQCGNKENTVALENGT	771		
y	1508	TAKKLG-----ETLTIKCGQTDHKNKLTDDNNIGVV--AGTDGFTVKLAKDLTN	1552		
b	772	TVKFLGNATFNGNTTIAA NSTLOIGNYVTDACVASADGTGTFVFEVNTGPITVTLNKQAP	831		

QY	1553	LNSV	-----NAG-----GTKID-----	-----DKGVSFVDSG--	1575
Db	832	VNAL	QITVSGPGNVVINEIAGNAGNHGAVTDTAFENSSLGAVELPKGIPFNDAQNTM	891	
QY	1576	--QAKANTPVL	SANGLLGGKVISNVCKGTKDTDAANVQOLNEVRNLLGLGNAGNDNA--	1631	
Db	892	PLTIKSTVGNKTA	KDFDVPSVVVLGV-----DSVIADQVIGDQNNIVGLG-LGSDNGII	945	
QY	1632	-----	DGNQVNIADIKKDPNSGSSNRVTIKACTVLGK-----	1665	
Db	946	VNATTLYAGISTLNN	QGTIVTLGSGVPMPTG-----TVYGLGTGIGASKFQVFTFTDYNN	1001	
QY	1666	-GN-----	NDTEKLATGGTGVGDKDG-----NANGD-----LSNVVVKTKQD	1702	
Db	1002	LGNLIATNATINDGV	TVTTGGI-AGIFDGGKITLGSVNGNVRFAFDGILSN--STSMI	1057	
QY	1703	GSKKALLATYN	AAQGTNYLTNNPBAIDRINEQGRIRFFHVNDGQE--PVVQGR-NGIDSS	1760	
Db	1058	GTRKA-----	NNGTIVLGN-----AFVGNIGSDTTPVASVFTGSDSG	1099	
QY	1761	AS-----	GKHSVAI-----GFOAKAD-----	1776	
Db	1097	AGLQNIYSQVIDP	GYNLGIVNSNIIILGGTTAINGKIDLVITWTLFASGTSTWGNNTS	1156	
QY	1777	-----	CEAAVAITGR-----TQAGNOSIAIGDNAQA--TGQDSIAIGTGNVVAG	1819	
Db	1157	IEVTLTLANGNIGH	IVILGSAQVNTTGTCTTIKQVDNANANFSGTQTYTLIOG---GAR	1213	
QY	1820	HSGAIGDP-----	STVRADN-SYSVGNNGNQFTDATQTDVF-----CVGN	1857	
Db	1214	FNGTLGPNFAVTG	SRNFVNSYLSIRANQDYYVITRNAENVVNDIANSPFGGAPGVQD	1273	
QY	1858	NIT--VTESNSVA-----	LGSNSATSA-----	1881	
Db	1274	NVTFVNTATNAY	NNLLAKNSANSANFVGAIVTDTSAITNVQLDLAKDIOAQLGNRL	1333	
QY	1882	GTOAKKSDGTAGT	TTTAGATGTVKGFGATGVAVSGVSGAERRIQNVAAG-EVSATST	1940	
Db	1334	G-ALRYLGT	PETAEMAG-----PEAGISAASAAGDEALDNVAYGIWAKPFYT	1380	
QY	1941	DA-----VNGSOLYKA-----	TQSIA-----ATNELDHRIHQENKANA-G	1976	
Db	1381	DAHOSKKGGLAGY	KAKTGTGVVIGLTDLNDNLNMGATGTTKIDIKHQDYKKGDKTDVNG	1440	
QY	1977	ISSAMAMASMPQ	AVIPGRSMVTGVIATHNQGGAVAVGLSKLSDNQWVFKINGSADTQGH	2036	
Db	1441	FSFSLYGA-----	QQLVKNFFA-----QGSATFSLNQVKNKSQRYE-----FDANGN	1482	
QY	2037	VGAAGVAGGFH	2046		
Db	1483	MSKQTAAGHY	1492		
RESULT	5				
OMPBL_RICHTY		STANDARD;	PRT;	1645	AA.
AC	P96989;				
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	01-OCT-2000	(Rel. 40, Last annotation update)			
DE	OUTER MEMBRANE	PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)			
DE	(SURFACE PROTEIN ANTIGEN)	(CELL SURFACE ANTIGEN) (SCAS) (ROMP B)			
DE	(CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)				
DE	(120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).				
GN	SLP.				
OS	Rickettsia typhi.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;				
OC	Rickettsiaceae; Rickettsia;				
OX	NCBI_TaxID=785;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=WILMINGTON;				
RX	MEDLINE=94040787; PubMed=8224886;				

RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
 RT "Cloning and sequence analysis of the gene encoding the crystalline
 RT surface layer protein of Rickettsia typhi.";
 RL Gene 133:129-133(1993).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN-ORIGIN: NCTC 10809;
 RX MEDLINE-92114896; PubMed-1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNR fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [3]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE-92104668; PubMed-1729180;
 RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent
 RT mutant deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 RN [4]
 RP STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L04661; AAB48987.1;
 KW Antigen; S-layer; Transmembrane; Cell wall.
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

 Query Match 4.4%; Score 457.5; DB 1; Length 1645;
 Best Local Similarity 21.7%; Pred. No. 1.3e-09;
 Matches 422; Conservative 206; Mismatches 634; Indels 683; Gaps 95;
 QY 234 SAYSLAVGLAATAEGOSTAIGS-----DATSSLSGALGAGTGAQLOQ-SIALGQGSVV 288
 DB 13 SAGLVASTATIVAGSGVAMGVMQYRNTTNAATVTDGAGFDGTCAGVNLPAVNSVI 72
 QY 289 TQSDNNSRPATYNTQALDPKFOATNNKAGPLSIGNSIKRKIIYVAGVKNKTDVAVNA 348
 DB 73 TANSNNAITENTNGN-LSLSLFDLTANTLA-VTINENTTLGFTNV----- 116
 QY 349 QLEAVVWKAKERRITFGDDNDSTVDKLGDLNTLTIKGAETNALTDDNNGVKEADNSGL 408
 DB 117 -----TKQG--NFFNTIGAGKSLITG-----HGITAQQAATTKSAQNVWS 156
 QY 409 KVKLAKTL--NNLTENVNTTLNATTTVKVSSSSSTTAELLSDSLTFQPTNGSQ----- 460
 DB 157 KVAGAAINDNLSGVGSIDFTAPSV-----LEFNLINPTTQEAPLTL 200
 QY 461 -STSKTVGVNGV-KFTN-----NAETAAIGTRITRDRKIGFARGDGVDEKQAPYLDKK 513
 DB 201 GDNAKIVNGANGILNITNGFVKVSKDTFAGIKTINI-GDNQGLMFNTTPDAANA-----L 254
 QY 514 QLKVGSAITIDNIDGADGNKKSINLAKSSANDAVT-----IE-----QLKAAKPTL 560

DB 255 NLOGGGNTINF-NGRDGTGKLVLSKNGNATEFNVTGSLGSLNKGVIETDFTAAGKLIA 313
 QY 561 NAGAGISVTPTE-----ISVDAKSGNVTAPTNTNIGVKTTELNSDGTSDFSYKGS 610
 DB 314 NGGAANAVIGTDNAGRAAGFIVSD--NGNAATISQVYAKDIVIOSANAGG----- 365
 QY 611 GTNNSLVTAEHL--ASYLNEVN-RTADSALQSFTVKEEDDDANAITVAKDTTKNA--GA 665
 DB 366 -----VTFEHLVDVGLGGKTNFKTADSKV-----IITENASEGS 399
 QY 666 VSLKLKAGKGLTVATKDKGTVPF-GLSQDSGLTIGKSTLNNDLGTVKDTNEQIQOVGANG 724
 DB 400 TDFGNLAVQ--IVVPNNKILITGNIGDAKNGNAGTAGVITFNANG-TLVSGNTDNIVVTN 456
 QY 725 IKFTNVGNSPGTGIANITARI--TRDKIGFAGS-----DGAVDTNPKYLDQDKLQ 772
 DB 457 IKAIEVEGA---GIVOLSGIHGAELRGNAGSIFKLADGTVINGPVNQ--PLVNNALA 511
 QY 773 VGNVKITNTGINAGKAITGLSPTLPSIADQSSRIELGN-----TIQDKDKSNAASIND 827
 DB 512 AGSIQDGSAL-----ITG-----DIGNGAVNAALQDITLANDAS--K 547
 QY 828 ILN-TGENL--KNNNPDIDFVSTYDIYDFANGNATTATVTHDTANKTSKVYVNVDDTT 884
 DB 548 ILTSGANIITGANAGAIHFQ-----ANGGTIQLTSTQ-----NNILVDFDLQVTT 593
 QY 885 IHLTGTDDNKKLGKVT--KLNKTSANGTATNFNVNSSDEDALVNAKDIENLNTL 939
 DB 594 -DQGVVDASLLNNQILTINGSTIGTICANTKTLGRFNVGSS--KTILNAGDVA--INEL 648
 QY 940 AKE-----IHTTKGPADTALQFTVKKVDENN-----NADDANAITVGOKNANNQVNT 987
 DB 649 VMENDGSVHLTH---NTYLITKTINAANOQKIIVAADPINTDTALADTGLGSAESPLSN 705
 QY 988 L--TLKGENGLINIKDKNGVTFGIN-TTSLKAKGKSTLNDGGLSINKPTQSEIQVGAD 1044
 DB 706 IHEATKAANGDSILHIGKGNLYANNITTDANVGLSFLHFRSGGTSTVSGTVG-----GQ 760
 QY 1045 GVPAKYNNGNVGAGIDGTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGGKKITNIQ 1104
 DB 761 GLKL-----NNLILD---NGTVKFLGDLITFNG-----GKTEGKSILQIS 798
 QY 1105 SGEIAQNSHDAVTGGKIYDLKTELENKISSAKTAQNSLHEFSVADEQGNNTFTVSNPYSS 1164
 DB 799 S-----NVITDHIESADN----- 811
 QY 1165 YDTSKTSDVITFAGENGITTKVNGGVVRGIDQTKGLTTPKLTVGNNGKG-----IVIDS 1220
 DB 812 -----TGTFLEFVNTDPTITLANKQAGYFGLVKQVMVSGPGNIAFNEITGNVAHAIAVDS 865
 QY 1221 QNGONTITG-----LSNLTANVTNDKGSVRTTEQGNIIKDEKTRAASIVDLVSAGFNQ 1275
 DB 866 ISFENASLGASLFLSGTGLDLVLTIKSTV----- 894
 QY 1276 GNGEAVDFVSTYDVTNPFADGNATTAKVYTDTSKTSKVYVDVNVDDTTIEYDKKLKGVKT 1335
 DB 895 GNGTVDNFENAPILVVSIGDSMNNGQV-----IGDQK----- 926
 QY 1336 TTLTSTGTGANKFALSNOATGADALVKASDIVAHLNTLSGDIOTAKAGASOANNSAGYVDAD 1395
 DB 927 -----NIIALS-----LGSDNSITVNSNTLYAGIRTTK----- 954
 QY 1396 GNKVIYDSTDNKYQAOKNDGTVDKTKVEAKDKLVAQAOTPDGTIAQNMVSKVINKEQVND 1455
 DB 955 -----TNQGTIV-----TLS----- 963
 QY 1456 ANKKQGINEDNAPYKGLEKAASDNKTKNAAVTV--GDLNVAQPTPLTFAGDTGTTAKKL 1513
 DB 964 ---GGIPNPGTIYGLGLENGDPKLVQVFTTDDYNLGSIIATNVTINDV----- 1011
 QY 1514 ETLTIKGGQTDNKNLTNNIGVAVAGTGTGVKLVAKDITNLNSVNAVAGTGTIDDKGVSFVDS 1573
 DB 1514 ETLTIKGGQTDNKNLTNNIGVAVAGTGTGVKLVAKDITNLNSVNAVAGTGTIDDKGVSFVDS 1573

Db 1012 -TLTTGG-----IAGTD-FDGKI-----TLGSING-----NANVKEVD 1043
 QY 1574 S-----GQAKNTPVLSANGLDLGGKVLSSNKGKTKDFAANVOQLNEVRNLLGLGN 1625
 Db 1044 TFSHTSMIVSTKANOVTY-----LGNALVGNL-----GSSDIPVASVR-----F 1085
 QY 1626 AGNDNADGNVNIADIKKDPNSGSSNRVIAKAGTVLGGKGNKTEKATGGIQGVGDKD 1605
 Db 1086 TONDSCVGLGNHNSONID---FGYNLTILNSDVLGG-----GTTAINGEIDL 1133
 QY 1686 GN-----ANGDLSNVVVKQDKSKALLATYNAAQOTNLTNNPAEADRINEQIRFEH 1741
 Db 1134 TNNLIPANG--TSTW-----GNNTSLSTLN-----1157
 QY 1742 VNDGNEPVVQCRNGIDSSASGKSHSVATGFOAKADCEAAVAIGROTOAGNSIAGDNAQ 1801
 Db 1158 VSGNVQIV-----IAEQAQVNA-----TTTGTITIKIQDNAN 1191
 QY 1802 A--TGQDSTAIAGTGNVAVAGKHGATGDPSTVKADNSYSVGNNGFT-----DATQTDVF 1853
 Db 1192 ANPSTGTQVTLIQG---GARFNGTLGAP-----NEDVTGNIEVKYELIRDANQDYVL 1241
 QY 1854 GVCNNITVTESNVALGNSAI--SAGTHAGTOAKKSDGCTAGTTTAGATGTVKGFAGQT 1911
 Db 1242 TRNDVNLNVVTVAV-----GNSAIANAPGVHONIAI-----CLESTDAAVNNML--LAKDS 1291
 QY 1912 AVGAVSYGASGABERRIONVAAGEVSATSDAVNGSOLYKATOSIANATNLDHRHONEN 1971
 Db 1292 SDVATEIGA-----IATDGAATVATVNLMDTO--KTQDLGNRLGAL--RYLSNSE 1338
 QY 1972 KANAGISSAMAMAPQA-----YIPGRSMVTGGIATHINGOGA--VAVGLSKL 2017
 Db 1339 TADVGGSETGAVSGGDDAIDQVSYGVWAKPFYINAEQDKKGLAGYKAKTAGVVGGLDPL 1398
 QY 2018 SDNQWVFKINGSADTQGHVGAAGV 2042
 Db 1399 ANDNLM-----IGAAIG 1410

RESULT 6
 ID OMPB_RICPR STANDARD; PRT: 1643 AA.
 AC Q53020; Q9ZCM0;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
 GN OMPB OR SPAP OR SPA OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=BREINL;
 RX MEDLINE=91045972; PubMed=2122457;
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RT "Characterization of the gene encoding the protective paracrystalline-
 RT surface-layer protein of Rickettsia prowazekii: presence of a
 RT truncated identical homolog in Rickettsia typhi,"
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BREINL;
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii,"
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E;

RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichertitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria,"
 RL Nature 396:133-140 (1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN=BREINL;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii,"
 RL Mol. Immunol. 29:95-105 (1992).
 RN [5]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RC MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: Identification of an avirulent mutant
 RT deficient in processing,"
 RL Infect. Immun. 60:159-165 (1992).
 CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; M37647; AAA26390.1; ALT_INIT.
 CC EMBL; AF161079; AAD42234.1; -.
 CC EMBL; AJ235273; CAAL5140.1; -.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREINL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TQEAPLTGA -> INSRSSYLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 7357DF392E6346CC CRC64;

Query Match 4.4%; Score 456; DB 1; Length 1643;
 Best Local Similarity 21.1%; Pred. No. 1.5e-09;
 Matches 399; Conservative 254; Mismatches 642; Indels 600; Gaps 101;
 QY 107 GGAAIAGSS---NKTVN--GSSLDKIGTADTGOESIAIGVDKASGDASIAIGSDDLHL 161
 Db 27 GFGVAMGAAMOYNRTTNAATFDGIFDQ-----AAGANIPVAPNSVITANAN--- 76
 QY 162 LDQHGNPKHPKGTLLINDLINGHVAFLKEIRSSKDNVDKYRRTTASGHASTAVGAMSYAQGH 221
 Db 77 -----NP-----ITFNPNGHLSNLSFLDTANDLAVTINEDTTLGFTITIA-----QQAQ 120
 QY 222 FSNAGFTRATKAYSGLAVGLATATAGCQTIAIGSDATSSLSGAIALGACTRAOQCSIA 281
 Db 121 FFN-----FTVAGKILNITGQGITVQESNTINAQNLTKVIG--GAINANDL 168

Db 827 VRVQARS-LADISLGAEGGATLGAVEAGSIDVRGGSTVAANSLSHANRD-----875
Qy 669 LKLGKNGLTVAIKDGTVTGSLSDSGLTIGKSTLNDGLTVDNBOIQVGANGIKFT 728
Db 876 VRVSGKDAYRVTAATSGG---GLHVSSGQLDGLGAVQARGALALD-----GGAGVALQ 925
Qy 729 NVNGSNPGTGIANATITRDKIGFAGSGADVNTKPYLDQDKLQVGNKYIINTGINAG--786
Db 926 SAKAS---GTLHVQGEHLDTGLAAVAGVDN-----GTGDRVAKLVSDAGAD 972
Qy 787 ---GKAIT-GLSTPLDSIAQDSRNIELGNTIQDKDKSNAASINDILATGENLKNNNPI 842
Db 973 LQAGRSMTLGIIVDTTDLQARQQKLEGLSVKSDGGLQAAA-----1013
Qy 843 DFVSTYDIVDFANGNATATVTHDTANKTSKVYVDVNVDDTTHILGTDDNKKLGKVTTK 902
Db 1014 -----CGALSIAAA-EVAGALELSQGVTVDRASARARIDSTSGVIGALK 1059
Qy 903 LNKTSANGNTATNFVNSDEDEALVNAKDIAEMNLTAKEIHTTKTADTALQTFVTKV 962
Db 1060 A-----GAVEAASPRRARRALRQDFTFGSVVVRAGNVTVGRG 1098
Qy 963 DENNADANAIVGOKNANNVNTLTLKGENGLNKTOKNGTFTVTEGINTTSGL---KAG 1019
Db 1099 DPHGVLAQGDIIIMDAK-----CTLLLR-----NDALTENGTVT--ISADSAVLEHSTI 1146
Qy 1020 KSTLNDGGLSIKNTGSEQIOVGADGVKFAK-----VNNNGVVGAG-----IDGT 1064
Db 1147 ESKISQSVLAAGDKGKPAV-----SVKVAKFLNGLTAVNDNNETMSGROIDVVDGR 1201
Qy 1065 TRITRDEIGFTGNGSLDKSPHLSKDGINAGGKKTINIOSGEIAONSHDAVTCGKIYDL 1124
Db 1202 POIT-----DAVTGEARKDESVDALVADGGPIV-VEAGELVHAGGIGNG-----1248
Qy 1125 KTELENKISSTAKTAAONLSHE-FSVADEQG-----NNFTV-SNPYSSTYTSKTSOV 1173
Db 1249 -RNKENGASVTVRTGNLVNKGYSACKQGVLEGGVGGALTNEFLVSGDGTQRIEQAOTENR 1307
Qy 1174 ITF-----AGENG-ITTKVNGVVRVIGIDOTGLTTPKLVGNNNGKGIVIDSQNGONTI 1227
Db 1308 GTFQSQAPAGTAGALVVKAAABAIHVDGVMAIKG---EMQIAGKGGSGPTVAGAKATTSA 1364
Qy 1228 TGLSLNLANVTN-----DKGSVRTT-----EOGNIKDEDKTRAASIVDVLSAGENLQ 1275
Db 1365 NKLSVDVASWDNAGSLDIKKGAAQVTVAGRYAEHGEVSIOGDYTVSA---DAIALAAQVT 1421
Qy 1276 GNGEAVDFVSYDT-----VNFADGNATTAKVTYDDTSKTSKVYVDVNVDDT 1322
Db 1422 ORGGAANLTSRHDTRFSNKRILMGPLOVAGGPPVSNGLKVRGVTVTAASFD---NET 1478
Qy 1323 TTEVKKLGVKTTTLTSTGTGAN--KFALSNOATGDALVKASDIIVAHNLPLSGDIQTAK 1380
Db 1479 GRAEVM-----AKSATLITTSGAARNAGKMQVKEAAT---IVAAS--VSNPGTFTA-----1522
Qy 1381 GASQANNAGYVDADG-----NKVIYDSTD-----NKYYOAKNDGTVDKTKEV-----AK 1425
Db 1523 GKDIITVTSRGGFDNEGKMSNKDVIKTEQFSNGRVLDAKHDHITVTSQADNRGSLKAG 1582
Qy 1426 DKLVAAQTPD--GTL-----AQMNVKSVINKEQVNDANKKOGINEDNAFVKGLEKASD 1478
Db 1583 HDFTVQAQRIDNSGTWAAAGHDATLAKAPHLRNTGQVVVAGHDHIIINSAKLNTGRVDAARD 1642
Qy 1479 NKTKNAAVT--VGDJLNAVAOTPLTFAGDT-----GT---TAKKLGTELTI 1518
Db 1643 IALDVADFTNTGSLYAEHDATLTAOGTORDLVDPQDHLPAVEGTLRVKAKSL--TTEI 1700
Qy 1519 KGGQ-----TDTNKLTDNNIGVVAGTD-----GFTVVKLAKDLT 1551
Db 1701 ETGNPSGLIAEVOENIDNKQAIIVGKDLTSSAHGNVANEANALLWAAGELTVK-AQNT 1759
Qy 1552 NLNS--VYAGGTTKIDKGVSVFVDSQGAQANTPVLISANGLDLGGKVISVNGKGTDXDAA 1609

Db 1760 NKRAALIEAGGNARLTAAVALLNKLGRIRA-----GEMHLDAPRIENTAKLSGEVQRK 1813
Qy 1610 NVQOL--NEVRNLLGLS-----NAGNDNADGNQVNIADIKKDPNSGS--SSNRRTVIKAG 1659
Db 1814 GVDVGGEGHGRSGGIGYVNYWLRAGNKKAG---TIA---APWYGGDLTAQESLIEVG 1866
Qy 1660 TVL---GKGKGNNDTEKLTATGGTQGVQVQDKNANGDLSN-----1694
Db 1867 KDLYLNAGARKDEHRLHLLNEGVIQAG--GHGHTGGVDNRSVVRTVVSANMEYFKTLPVSL 1924
Qy 1695 -----VW-----VKTKD-----1702
Db 1925 TALDNRAGLSPATWNFQSTYELLDYLDQNRXEYIINGLYPTTYSVNTLKNLDLGYQAK 1984
Qy 1703 -----GSKALLATYNAAGOTNYLTNNPAEIDRI 1732
Db 1985 PAPTAPMPKAPELDLRGHTLESAGRKIFGEYKKGLOGEYKAKMAVQAVEATRRV 2044
Qy 1733 NEQ-GIRFFHVNDQNPVQGRNGCIDSSASGKHSVAI--GFOAKADGEA-----1779
Db 2045 HDLQGRYGRKALGMDAETKEVDGIIQEFADLRTVYAKQADQATIDETDKVAQRKVSQ 2104
Qy 1780 -----AVAIQROTQAGNOSIAIGDNAQATGDQSIAGTGNVYVAGKHSIGAIDPSTVKA 1832
Db 2105 IDAVRLQAIQGRVTLAKALSAAALGADWRALGHSOLMQRWKDEKAGKRGAEIA-----2157
Qy 1833 DNSYSVGNNOFTQATQDQDFGVGNNTVTESNVALGSNSAISAGTHAGTO--AKKSDG 1890
Db 2158 -----FYKPEQT-VLAAGAGETLS-NGAIHNGENNAQNRGPEGLKIGHATS 2204
Qy 1891 TAGTTTATAGTGVKGFAGQTAAGVAVSG-----ASCAERRIONVAAGEV-----SAT 1938
Db 2205 VSGSDALRDVLEKRLDIDDLAALVNPHTFRIGAAQTSADGAGPALARQARQAP 2264
Qy 1939 STDV-----NGSOLYKATQSTANTNDELHRIQENKANKAGIS--SAMAMASMPQAYI 1991
Db 2265 ETDMGVMDARGLSGSALASLASLAAOGLSVGRRAQVADAGLAPSAVAAPVAGADV 2324
Qy 1992 PGRSMVTGGIATHNGGAVAVGLSK 2016
Db 2325 -GVPEVTG---DQVDQPPVAVGLEQ 2345

RESULT 9
P2P_LACPA STANDARD; PRT; 1902 AA.
ID Q02470;
AC 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE) (LP151).
GN PRTP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1597;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 151;
RX MEDLINE=92381481; PubMed=1512565;
RA Holck A., Naes H.;
RT "Cloning, sequencing and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151.";
RL J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
RP SEQUENCE OF 189-196.
RX MEDLINE=92226694; PubMed=1564442;
RA Naes H., Nissen-Meyer J.;
RT "Purification and N-terminal amino acid sequence determination of the
RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.
RT paracasei.";

RL J. Gen. Microbiol. 138:313-318(1992).

CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE

CC GROWTH OF THE BACTERIA ON MILK.

CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD

CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED.

CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND

CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS, AND

CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED

CC INSULIN B-CHAIN.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

CC SUBTILASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M83946; AAA25248.1; --

CC PIR; B44858; B44858.

CC HSP; Q99405; IMPT.

CC MEROPS; S08.019; --

CC InterPro; IPR000209; --

CC InterPro; IPR001899; --

CC Pfam; PF00746; Gram_pos_anchor; 1.

CC Pfam; PF00082; Peptidase_S8; 3.

CC PRINTS; PR00723; SUBTILISIN.

CC PROSITE; PS00136; SUBTILASE_ASP; 1.

CC PROSITE; PS00137; SUBTILASE_HIS; 1.

CC PROSITE; PS00138; SUBTILASE_SER; 1.

CC PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.

CC Hydrolase; Serine protease; Cell wall; zymogen; Signal;

CC Transmembrane. 1

CC SIGNAL 33 POTENTIAL.

CC PROPEP 34 187 POTENTIAL.

CC CHAIN 188 1902 PII-TYPE PROTEINASE.

CC DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).

CC TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).

CC FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).

CC FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE

CC FT PROTEINS.

CC FT SEQUENCE 1902 AA; 200253 MW; D8C9F38CEE5DA582 CRC64;

Query Match 4.3%; Score 447.5; DB 1; Length 1902;

Best Local Similarity 20.7%; Pred. No. 3.5e-09;

Matches 460; Conservative 237; Mismatches 831; Indels 689; Gaps 98;

QY 40 SVCTLSFARIAALAVLYGATLGSAYAAKDKTKHIALGEQNPQRSTKADGDRAIAI 99

DB 9 SILLAGTVAGALAVLPVGEIQAKAAISQOTKVSSLA-----NTVKAATAKQ-----AA 57

QY 100 GENAAGCGQAATGSSNKTNGVSSLDKI-----GTDATGOES- 137

DB 58 TDTAATNTQAIAQLAKGIDYKLNKVVQOQDTYVDVIVOMSAAPASENTLRDYSST 117

QY 138 -----TAIGDVKAA-----SGDA-----SIAIGSDHLHLDQHGPK- 169

DB 118 AEIOQETNKVIAAQAQAAVEQVQTQTAGESYGVVNGFSTKVRVVDIPKLQIAGVK 177

QY 170 -----HPKGTLLNDLINGHVL-----KEIRSSKDNVVKY 199

DB 178 VTLAKVYPTDAKANSMAVQAVNSNYKYKGGTWSVIDTGDPTHKDMRLSDDKVKL 237

QY 200 RR-----TTASGH-----ASTAVGAMSQAQ-----HPSNFGTRATAKS 234

DB 238 TKYDVEKFTDAKGRVFTSKVPYGFVNDNDITDITDVTDEQHGHHVAGIANGTGDD 297

QY 235 AYSLAVGLAATAEGOSTIAIGSDATSSSLGATGAGTGAOLQGSIALGOGSVVTSQDNN 294

DB 298 PTKSVGVGAPEAQLAMKVFTNSDTSATTSATLSVA-----IEDSAKICADV-----NM 348

QY 295 SRPAYTPNTQALDKFQATNNTK-----AGPLSIGSNSIKRKLIINVGAGVKNKTDAVNAQL 350

DB 349 SLGSDSGNQTLEDPEIAAIVQANANESGTAIVISAGNSGTSG-----SATQGVNK-DYYGLQDN 404

QY 351 EAVYKWKAKERRITFGDDNSTDVKIGLNDTLT-----IKGGAETNALTDNN----- 396

DB 405 EMVGTGTSRGATTVASAENTDV-ISOAVTITDGKLDLQGLPETIQLSSNDFTGSDQKFF 463

QY 397 -----IGVVKEAD-NSGLKVKLAKTL-----NNUTE 421

DB 464 YVWKDASGDLKGAADYADAKGKIAIVKRGELNFADKQYAAQAAGLIIVNNDGTA 523

QY 422 VNTITLNNATTVKVGSSSTTAEILSDSLTF-----TQPNTGSGSTSKT 465

DB 524 TPLTSIRLTITFTPTFGUSSKTGQKLVDMWTAHPDDSLGKIALTLPLPNQKYTEDKMSDT 583

QY 466 VYG-VNGVKFTNNAETTAAGTTRITROKIGFARDGDVDEKQAPYLDKKQKLVGSAITI 524

DB 584 SYGVSNLSP--KPDITAPGGINSTQNNNGYTNMSGT-SMASPFIAGSQALLKQALNKK 640

QY 525 DNGIDAGNKKISNLAKGSANDAVTIEQLKAAKP-----TLNAGAGISVTPT 572

DB 641 NNPFVADYKQL-----KGFTALTDFLKTVMETAQPIINDINYNVIVSPRROGAGL--VDVK 694

QY 573 ISVDAKSNVTAPYINIGVKTTELNSDGTSKFSVKGSGTNSLVTAEHLASLYNEVRT 632

DB 695 AAIDALEKNPSTVVAENGYPVAVELKDEFTSDK-----TFKLFTTNR 736

QY 633 ADSALQSFVKEEDDDANATVAAKDTTKNAGAVSILKLGK-----NGLTVATKKDQTV 687

DB 737 T-----HELTQYQMSNTDITNAVYTS-TDPNSGLVYDKKIDGAAIKAGSDIIVPACKTAQI 791

QY 688 TFGLSQDSGLTIGKSTLNNDGLTVKDTNEQIQGVANGIKFTNVNNGSPGTGIANTARITR 747

DB 792 EFTLS-----LPKSFDDQQFVEG-----FLNFKGS-----DGSRLNL 823

QY 748 DKIGFAG--SDGAVDTNPKPYLDQDKLVGNVKITNGINAGGKAITGLSLPSTADOSS 805

DB 824 PYMGFFGMDNGKI-----VD-----SLNGITYSPAGNYG-----TVPLLTNKT 864

QY 806 RNIELGNTIQDKD-----KNSAASINDILFTGNLKNNNNPIDFVSVDYD 852

DB 865 GHQYGGVMTDADGKQTVDDQAIAPSSDKNALYNDISMQYILLRNISN-----VQVDILD 919

QY 853 FANGN-----ATTATVTHDTANKTSKVYDVNVVDTTIHLTGTDNNKKGKLVKTKLNK 905

DB 920 -GQGNKVFTLSSSTNOTKTYVDAHSQKYIYNAPAWDGTYY-----DQDGNIK----- 967

QY 906 TSANGTATNPNVNSDDEDALVNAKDIAENLTLAKEIHTTKGTADTALQFTVVKVDEN 965

DB 968 -----TADGSGTYRISGVPEG 984

QY 966 NNAD-----DANAITVGOKNANNQVNTILTKENG-----LNTKTDKNG----- 1004

DB 985 GDKQVDFVDFPKLDSKAPTVRHV-----ALSATENGKTQYIYLTAEAKDDLGLDATK 1037

QY 1005 TVTFGINTTSLKA---GKSTLNDGLSIKNPTGSEQIQVGADGVKFAKNNVNGVVGAGI 1061

DB 1038 SVKTAINEVNTLADATFTDAGTYTKIETPLSDEQAQALNG----- 1081

QY 1062 DGTTRITRDEIGTCTNGSLDKSKPHLSKDG--INAGG--KKITNIQSGEIAQNSHDAVT 1117

DB 1082 DNSAELYLTDNASNATNQDASVQKPGSTFSLIVNGGGIPDKISSTTTG-----YEANT 1135

QY 1118 -GGKTYDLKTELENKISSTAKTAQNSHEFSVA-DEQGNFTVSNPYSYDTSKTSVDIT 1175

DB 1136 QGGTYTTSCTYPAADVGTYTDAGQKKHDLNTYDAATNSFTASMAVNTADYAAQVDLYA 1195

QY 1176 FAGENCI---TTKVNKG---VRVGIQDTKGLT-----TPKLTVCNNNGK 1214
Db 1196 DKAHTQLLKHFDTKVRLTAPTFTDLKFNNGSDQTSATIKVTGTVSSDTKTWNVGDVTA- 1254
QY 1215 GIVIDSON-----GQNTITGLSNTLANTVNDKSGVRRTEQNIKIDEDKTRAASI 1264
Db 1255 --ALDAQHHFSDVPVNYGDNIIK-----VATDEDDGNT-TTEQKTISSYDP----- 1299
QY 1265 VDLSAGFNLOQNGEAVDFVSYDT-VNFA--DGNATTAKVYDDTSKTSKVYDVNDVDD 1321
Db 1300 -DVLK-----NAV-----TFDOGVKFCANEFNATSAKF-YDPKGTGATITGVKHP 1344
QY 1322 TTIEVDKDKLVK-----TTTLSTGTGANKFAL-----SNQATGDALVKASDVAHLN 1370
Db 1345 TTLQVQDKQISIKNDLTFSTLDTGLTGOKPGVVGVVGDITQNKTFQEAFLTILDAVA--P 1402
QY 1371 TSLSDIOTAKGASQANNSAGYVDADGNKYIYDSTDKYQAKNDGVPDKTEKVAKDKLVA 1430
Db 1403 TSLSDST-----DAPVYNDPNFQITGTATDNAQVLS-----LAI 1438
QY 1431 QAOPPDGTLAOMNVKSVINKEQVNDANK--QGINEDNAFVKLEKASDNKTKNAAVTV 1488
Db 1439 NGSVASQYADININSKPGHMAIDQPVKLEGN-----VLTVAVTDSENNTTKKIIV 1493
QY 1489 GDNLVNAQTPLTEAGDTGTTAKKLGETLTIKGGTQDTNKLTDNNIGVAGTGDGFTVKLAK 1548
Db 1494 -----YEPKKTLLAIPVTPS-----TTEPAKTVI--LTAN-----AAATGETVOYSA 1534
QY 1549 DLTNLNSVNAAGTKIDD--KGVSFVDSGQAKANTPVLNSANGLDGGKVISNVGKGTGD 1605
Db 1535 D-----GGKTYQDVPVPAAGVT-VTANGTFKPSDLYGNESPAVDYVVTNI-----KA 1580
QY 1606 TDAANVOQLNEVRNLGLIGNAGNDNADGNVNIADIKKDPNSGSSSRNRTVIKAGTVLGGK 1665
Db 1581 DDPACQLQTAQALNLT-----IASAKTSLASCKYDDATT-----TALAAA 1620
QY 1666 GNNDEKLTAGGIGQ--VGVDRKDNANGDLSNWKTKQDQSKKALLATYNAAGQTNLTN 1723
Db 1621 ----TQKAQTAQDQDASVDSLTGANRLQD-----AINQLAAKLPAKDKTSL-- 1665
QY 1724 NPAAEIDRINEOGIRFPHVNDGNPPVQGRNGIDSSASGRKHSVAIGFOAKADGEAAVAI 1783
Db 1666 -----NQLOSVAALGTDLGNQDTPSTGKFTTAAALDLVA--- 1700
QY 1784 GRQTAQNGQSTAGNDAQATQDQSTAGTGNVACKHSGAIGDPSTVKAADNSVGVNNO 1843
Db 1701 --QAQATQGT--ADQLQA-----SLAKVLDVAVLAKLAEG-----IKAAATPAEVGNAK- 1743
QY 1844 FTDATQTDVFGVGNITVTESNSVALGSN-----SAISAGTHAGTQAKKSDGT 1891
Db 1744 --DAATGKTWYADIADTLTSCQASADASDKLAHLQAQLQSLTKVAVAEAAKTAGKGDGT 1801
QY 1892 AGTITTAGATGTGVKGFAGQATVAGVSGVSGAERKIONVAAGEV-----SATSTD 1941
Db 1802 TGTSDKGGGGTTPAPAGDT-----GKDKGDEGQSPSSGGINPTKPTATTSTSTD 1851

RESULT 10
OMPBL_RICJA STANDARD; PRT: 1656 AA.
AC OMP653:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE [SURFACE PROTEIN ANTIGEN] (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=35790;

RN SEQUENCE FROM N.A.
RP STRAIN=VH;
RC Uchiyama T.;
RA "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB003681; BAA20138.1; .
KW Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1338 32 KDA BETA PEPTIDE.
FT CHAIN 1339 1656 POLY-GLY.
FT DOMAIN 528 533
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;
SQ

Query Match 4.3%; Score 443.5; DB 1; Length 1656;
Best local similarity 21.4%; Pred. No. 4.1e-09;
Matches 439; Conservative 237; Mismatches 750; Indels 623; Gaps 106;

QY 191 SSKDNDVKYRRRTASGHASTAVGAMSYAQGHFSNAGFGRATKASVLAUAGLAETAEQS 250
Db 2 AOKPNFLKLLISAGLVSTASTATIVASFAGSAGMAAIOQNRTN-----GVATTVDG-- 52
QY 251 TTAIGSDATSSSLGALGACAGTGAQLOGSIALGCGSVVVTOSDNNRKPATPTNQALDPKF 310
Db 53 ---VGEDQT-----VAL-----ANVAVAPNAVITANANG----- 79
QY 311 QATNNTKAGPLSGISGSIKRIKIIINVAGVYKNTDANVVAOLEAVKWKERRITF--QDSD 368
Db 80 -INLNTPAGSEN-----GLFLSNANLA-----VTVSEDTTLGFINNAAN 118
QY 369 NSTDVVIGLD--NTLTIKGAETN---ALTDNNIGVYKADNSGLKVLAKTLNLTENV 423
Db 119 NANRFLTLIDACKTLTITGOGITNVQSAATHNAQNLIVAKF--NGGAAI-----ANNDLSGLG 173
QY 424 TTTLNATTTVKVSSSSSTTAE-----LLSDSLTFTQPTNG-----SOSTSKTVYG 468
Db 174 TIDFGAASTLVFDLANPTQKAPLILADNALIVNGANGTLNVNTNGFIQVSKSEATVKA 233
QY 469 VN-----GVKFTNNAETTAAGITRTITROKIGFARGDGDVDEKQAPYLDKKKLVGSAIT 523
Db 234 INIGDGGGFMFTNATNANAL-----NQAGGTTIN 264
QY 524 IDNGIDAGNKKISNLAKGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISYDAKSGNTV 583
Db 265 F-NGTD--GTGRVLVLLSKNGAATD-----FNVTSGLGNLKGII-IELNTVAINGQLI 312
QY 584 APTYNIGVKYTELNSDGTSDKFSVKGSGTNNLSLVAHLASYLNEVNRNTRADSALQSFYVK 643
Db 313 A---NAGPANAVI-----GTNNG---AGRAAGFVVS----- 338
QY 644 EEDDDANATV-----AKDTTKNAGAVSLKLGKNGLT-----VATKKDGTWTF-- 689
Db 339 ----DNKKAATIDQGVYAKD-----WVQSANANGOVNFRHIVDVGIDGTAFKTA 386
QY 690 ---GLSQDSGLTIGKSTLNN--DGLTVKDTNEQIQVGANGIKFTNVNGNSPGTGIANTAR 744

Db 397 SIVAITONS--NFGTDFGNLAAQVTPVDT---MTLTCN---FTG-DANNPG-----NTAG 433
Qy 745 ITRDKIGFAGSDGAVDTNPKPYLDQDKLQVGNVKTN--TGINAGGKAITGLSTPLPSIAD 802
Db 434 V-----ITFAANGTLASASAD-----ANVAVTNNTALTEASGVGVQLSGT-----474
Qy 803 QSSRNIELGNT-----TODKDKSNAASINDILNTGNFKNNNNPIDFVSTYDVIDVFANGN 857
Db 475 -HTAELRLGNAGSVFKLADGTVIN-GKVNQTVLVGGVLAAGAITLDGSAIT-TGDIGNGG 531
Qy 858 ATTATVTHDTANKTSKVYVDVNDVDTTILHTGTDNDKKGKLVKTKLAKTSANGNTATNFN 917
Db 532 CGAALQSTLANDATK-----TLTLGGA-----NIIISANGGT-INFQ 567
Qy 918 VNSDEDAVNNAKDIAENLTLAKEIHTTKGTADTALQTFVTKVDENNNADANAITVG 977
Db 568 ANG-----CTIKLITSPQNNIVVD-----CDLAIAT-----DOTGVVDASSLTNA 606
Qy 978 OKNANNQVNTLTLKGENGLNKTDKNGVTPEGI-----NTTSG-----LKAGKSTLNDGGLSI 1030
Db 607 Q-----TLTISG-----TIGLIICANNITLQGFNIGSKTTLNGGNVAI 644
Qy 1031 KNPTGSEIOVGAGDS-VKFAK-----VNNNGVVCAGID-GTTR 1066
Db 645 -----NELVIGNNGSVQFAHTYLTITRTNAAGOKIIFPNPVNNNTTLAAGTNLGSAA 698
Qy 1067 ITRDEIGFTGNG-----SLDKSPHLSKDGINAGKKTITNTQSGEI 1108
Db 699 NPLAEINF-GSKGARADTVLNVGEGVNLATNIITDANVGSPFVFNAGK---NIVSGTV 754
Qy 1109 AQNSHDVATGKGIYDLKTELENKISSAKTAQNSLHEFSVADEOGNNTVSNPSSYDTS 1168
Db 755 -----GG-----QQCKENTVALDNGTTVKFELGNATFNNGNTTIAANSTLQISGN 798
Qy 1169 KTSVDITPAGENGITTKVKNQVVRGID-OPKGLTTPKLTGVNNGKGIIVDSONGQNTI 1227
Db 799 YTADEFASADGTGIVEFVNTGPIWLNKQAPVNALQIIVTSGFNGVNVNIEIGNAGNYH 858
Qy 1228 TGLSNTLANVTDNRKSVRTTEBQNIIDKEDTRAASIVDVLSAGFNLOGGEAVDF-----1283
Db 859 GAMTDITAFENSSILGAVLFLPSPGPFENDAGNTIPLTKSTV-----GNETAEGFSVPS 911
Qy 1284 -VSTYDITVNFADG-----NATTAKVYDDTSKTSKVYVDVNV 1319
Db 912 VIVSGVDSV-IADGQVIGDQNNIVGLGLSDNGIIVNATTLIYAG-----IGTINN 960
Qy 1320 DDTTIEVKDKKLGKVTTLT-----STGFGANKFALSNOATGDALVKASDIVAHLNLTSGD 1375
Db 961 NQGVTVTLGSG---GVNPTGTVYGLGTGIGASKFKQVFTTD--YNNLGNIIATNTTINDG 1015
Qy 1376 IQTAKGASQANNSAGYVDADG-----NKVIYDSTDNKYQAKNDGTVDK 1419
Db 1016 VTVTTGGIAGGIAG-TDFDCKITLGSVNGNANVRFADGIFSNSTSMIVTKKANNCTVY 1074
Qy 1420 TKEVAKDKLVAQAOTP-----DGTLAQMNVKSVINKQVNDANKKQGINEDNAFVKGLE 1473
Db 1075 LGNAFVGN-IGDSOTPVASVRFSTGNSNAGLKNYISQVIDFGTYNLGIVNSVILGGST 1133
Qy 1474 KAASDNKTNAATVVDLNAVAQTPPLTFAGDTGTAK--KLGETILTIGKQTDINKLTDN 1531
Db 1134 TAIN-----GKIDLLTNT-LTFAGGTSTWGNNTSTETTLTLANG-----1171
Qy 1532 NIGVAGTGDGFTVKLAKDLTLNLNSVNAAGTKI---DKGVSVFVDS-----SGOAKANTP 1582
Db 1172 NIGHIVIAEG-----AQVNATTGTTINIVQDNANANFSGTQTYTLLOGGARENGT 1222
Qy 1583 VLSANGLDLGKVTSNVKGKTKDTDAAN-----VOOLNEVRNLL--GLGNAGNDNADGNQV 1636
Db 1223 LGGFNFTVTSNRNFVNYGL-----IRAAQNDYVITRTNAENAVINVDINTISFPFGAPGVQ 1278
Qy 1637 NIADIKKDPNSGSSNRNVIK-----AGTVLGGKGNNDTEKLTATGCIQGVGDKD---1685

Db 1279 NVTFVNTATNTAAVNNLLAKNSADSANFVGTV-----TDTSAAITNA-QLDVAKDIOA 1332
Qy 1686 --GNAGDLSNVWVKTKQDGSKALLATYNAAQGTNYLTNNPAEAIIDRINEOGI---REFH 1741
Db 1333 QLGNRLGALRYLGT-PEMVGSEAGAI PAAVAAGD-----EAVDNV-AYGIWAKPFP 1381
Qy 1742 VNDNQEPVVOGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGROTQAGNQ---SIAIGD 1798
Db 1382 T-----DAHQSCKGGLA-GYKAKTTG--IVIGLDTLANNLMIGAIGI 1422
Qy 1799 NAQTGQDSQIAIGTGNVVAGKHSIGAIDPSTVKADNSYSVGN---NOFTDATOTDVF 1854
Db 1423 TKTDIKHODYKKDKTDVNGFSFSLYGAQFV--ENFFAQGSAIFSLQVKNKNSRYEFD 1480
Qy 1855 VGNITVESNSVALGNSAISAGTH-----AG-TQAKKSDCTAGTT 1895
Db 1481 ANGNM---SKQIAAGNYDNMTFGNLTGVGDYNAMQGVLTPTMAGLSVLKSSDENYRET 1536
Qy 1896 TTAGATGTVKG-FAGQTVAGVSGVSGAERRIQNVAAGEVSATS---TDAVNGSQLYKA 1951
Db 1537 GTTVANKQVNSKFSRDT---DLIVGA-----KVAGGTMTNITDFAVYPEVHAFVHKV 1585
Qy 1952 TOSTANATNELDHR-----IHONENKA---NAGISSAMAMASMPQAYIPGRSMVTGGIAT 2003
Db 1586 TGRUSKTQSVLDGGVTPCISQPDRTAKTSYNLGLSASIRSDAKMEYIGYDAQIASKYTA 1645
Qy 2004 HNGOGAVAV 2012
Db 1646 HOGTLKVRV 1654

RESULT 11
PIP_LACLC
ID PIP_LACLC STANDARD; PRG: 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE
DE DE PROTEINASE).
GN PRTP.
OG Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Plasmid pW05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WG2;
RX MEDLINE=88149035; PubMed=3278687;
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris Wg2".
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -!- GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSTRATE PREFERENCE HAVE BEEN NOTED.
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Thu Sep 13 14:18:59 2001

```
CC -----
DR EMBL: M24767; AAAL7677.1; -.
DR HSSP: Q99405; 1MPT.
DR MEROPS: S08_019; -.
DR InterPro: IPR000209; -.
DR pfam: PF00746; Gram_pos_anchor; 1.
DR pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1..
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
DR Hydrolase: Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 4.3%; Score 441.5; DB 1; Length 1902;
Best Local Similarity 20.7%; Pred. No. 5.7e-09;
Matches 459; Conservative 238; Mismatches 829; Indels 693; Gaps 97;

QY 40 SVCTLSFARIAALAVLIGATLCSAYAAQKDKTHIAIGEQNPRRSGTAKADGDAIAI 99
DB 9 SILLAGTVAGCALAVLPVGEIOAKAAISQQTGGSLA-----NTVTPATAKQ-----AA 57
QY 100 GENAQAQGAATGSSNKTNGSSLDKI-----GTPATGOES- 137
DB 58 TDTTAATTQAALQALAAKGIDYNKLNKVKQOQDIYDVIVQMSAAPASENGTLRTDYSST 117
QY 138 -----TAIGDVKVA-----SGDA-----SIAIGSDDLHLHDQGNPK- 169
DB 118 ABIEQETNKVIAAQAASVKAAYEQVTOQTAGESYGVVNGFSTKRVVDIPKLAQIAGVKT 177
QY 170 -----HPKGLINDLINGHAVL-----KEIRSSKDNVDKY 199
DB 178 VTLAKVYPTDAAKMANVQAVMSNYKYKGEQTVSVSDIGDPTHKDMRLSDDKDKVL 237
QY 200 RRTTASGHASTAVG-----AMSYAQG-----HFSNAFGTRATAKS 234
DB 238 TKSDVEKFTDTAKHGRYFNFSKVPYGFENYADNNDITDITDVEQHGHHVAGIIGANGTGDD 297
QY 235 AVSLAVGLAATAEGOSTAIGSDATSSSIALGALGAGTQAOLGCSIALGOGSVVYQSDNN 294
DB 298 PAKSVGVGAPEAQLLAKMVFYNSDTSATYGSSTLVA-----IEDSAKICADV-----NM 348
QY 295 SRPAPTPTQALDPKFPQATNNTK-----AGPLSIGNSIKRKLIINVGAGVKNKTDAVNAQL 350
DB 349 SLGSDSGNQTLEDPELAAYQONANESGTAAVISAGNSGTSG-----SATEGVNK--DIYGLQDN 404
QY 351 EAVVKWAKERRITFOGDDNSTDV-----KIGLDNLTILKGAETNALTDNNI 397
DB 405 EMVGTGCTRGATTVASAEVDITQAVTITDGTGLQLG-PGTIQLSSNDFGSPDQKFF 463
QY 398 GVVKE-----ADNSGLVKVLAK-----TLNLT 420
DB 464 YVVKDASGNLSKALADYTADAKG-KIAIVKRGELSFDDKQYAAAGAAAGLIIVNNDGT 522
QY 421 EVNTTILNATTIVKVGSSSSTTAEILSD-----SLTFTQPTNGSQSTSK 464
DB 523 ATPVTSNALTTFTTFCGLSVTGCKQLVDVWTAHPDDSLGLVKIALTLVPNQYTEDKMSDF 582
```

```
QY 465 TVYGV-VNGVKFTNNAETTAAGTTRITRDKIGFARDGDVDEKQAPYLDKKQLKVGVSVAIT 523
DB 583 TSYGPPVSNLSE--KPDITAPGGNIWSTONNGNNGSGT-SMASPFIAQSQAALLKQALNN 639
QY 524 IDNGIDAGNKKISNLAKGSSANDAVTIEQLKAAKP-----TLNAGAGISVTP 571
DB 640 KNPFPYAYKQL-----KGTALTDFLKTVMEMTAOPINDINYNVIVSPRQAGL--VDV 693
QY 572 EISVDAKSGNVAPTNYNIGVKTTELNSDGTSDKFSVKGSGTNNSLVTAEHLASYLEVNR 631
DB 694 KAAIDALEKNPSTVAENGYPAVELKDFSTDK-TFKLTFTN-----NGLTVATKKDGT 666
QY 632 TADSALOSFTVKEEDDDANAITVAKOTTNAGAVSLKLGK-----NGLTVATKKDGT 666
DB 735 ---STTHELTYQMDSNITDNAVYTSA-TDPNSGVLYDKKIDGAIAKAGSNITVPAGTAQ 790
QY 687 VTEGLSQDSGLTICKSTLNNDLGVKDTNEQIQVANGIKFTNVNNGSGTGTANTARIT 746
DB 791 IEFITLS-----LPKSPDOQFVGS-----DGSRLN 822
QY 747 RDKIGFAG--SDGAVDTNKPYLQDQKLVQGVNKTNTNGINAGKAITGLSPTLPSIADQS 804
DB 823 LPYMGFFGDWMDGKI-----VD-----SLNGITYSPAGNFG-----TVPLLTKNK 863
QY 805 SRNIELGNTIQDKDKSNA-----ASINDILNTGFNLKNNNNPDIQFVSTYDIV 851
DB 864 TGTQYGGWTDADGNQTVDDQATAFSSDKNALYNDISMKYLLRLNISN-----VQVDIL 918
QY 852 DPANGN-----ATTATVTHDTANKTSKVYDVNVDDTTIHLTGTDNKKLVGKVTTKLN 904
DB 919 D-GQGNKVTTLSSSNLTKTYYNAHSQYIYVYVAPAMWDTYV-----DQDGNK- 967
QY 905 KTSANGNTATNPNVNSSDEDALVNAKDAENLTLAKEIHTTKGTADTALOTFTVKKVDE 964
DB 968 -----TADDSGYTYRISGYPE 983
QY 965 NNNAD-----DANAITVGQKNANNQVNTLTLLKGNG-----LNITDKNG----- 1004
DB 984 GGDKRQVDVFPKLSDKAPTVRHV-----ALSAKTENGKTQYLYLTAEAKDDLGLDALT 1036
QY 1005 -TVTFGINTSGLKA---CKSTLNDGGLSIKNPTGSEQIQVGCADGVKFAKYNNGNVGAG 1060
DB 1037 KSVKTAINEVTNLDAFTDAGTADGYTKIETPLSDEQAQALGNDGNSAEL----- 1087
QY 1061 IDGTTTRITDEIGFTGNGSLDKSPHLSKDGINAGG--KKITNIQSGEIAQNSHDVIT- 1117
DB 1088 -----YLTDNASNATDODASVQKPGSTFSLVNGGGIPDKISSTTG-----YEANTQ 1136
QY 1118 GGIYDLKTELENKISSHTAKQNSLHEFSVA-DEOGNFTVSNPYSSYDTSKTSDVITF 1176
DB 1137 GGGTYTFSGTYPAAVDGTYTNAQKKHIDLNTTYDAATNSFTASMPVYNADYAAQVDLYAD 1196
QY 1177 AGENGI-----TTKVNKGV-----VRVGIDQTKGLT-----TPKLTGVNNGNGK 1215
DB 1197 KAHTQLLKHFDTKVRLTAPTFTDLKFNNGSDQTSATIKVTGTVSADTKTVNVGDTPVA-- 1254
QY 1216 IVIDSQN-----GQNTITGLNLTANVTNDRKSGVTTTEOGNIKDEKTRAASIV 1265
DB 1255 -ALDAQHFHSVDVPVNYGDNITIK-----VTATDEGNT-TTEQKITSYD- 1299
QY 1266 DVLSAGFNLOGNEAVDFVSTYDTVNFADGNATTAQVYDDTSTSKVYDVNVDDTTIE 1325
DB 1300 DMLK-----NSVTFOGVTFGANEF---NATSAKF-YDPKGTGIATITGKVKHPTTLQ 1348
QY 1326 VKDKKLGVK-----TTTLTSTGTGANKFAL-----SNQATGALYKASDIVAHLTSLG 1374
DB 1349 VDGKQIPIKDDLTFSFTLDLGTGLQKPFVVVGDTTQNKTFQEQALFTILDAVA--PTLSL 1406
QY 1375 DIQTAKGASQANNAGYVDADGNKVIYDSTDN-KYYQAKNDGT-----VDKTRKVAOK 1427
DB 1407 ESST-----DAPVTNDPNFOITGTATDNQAYLSLSINGSSVASQYVDININSKPG 1458
QY 1428 LVAQAQTPDGTLAQMNKSVINKEDQVNDANKKQGINEDNFAVKGLEKAASDNKTKNAVT 1487
```


Db 1459 HMAIDQ-PVKLEGNLVTV-----AVTD-----SEDTTKNITVYEPKTL-AAPT 1505
Qy 1488 VGLNVAQPTPLFAGDTGTTAKKLGELTITKGGQDTNKLTDNNGVAGTGTGFKLA 1547
Db 1506 V-----TP-----SITEPAKTVLTANSAT-----GETVOYS 1533
Qy 1548 KDLNLNSVNAAGTKIDD---KGVSVFVSSGQAKANTPVLSANGLDLGGKVISWVGKTK 1604
Db 1534 AD-----GKTYQDVPAAGVT-VTANGTFKFKSTDLYGNSPVDVYVVTNI-----K 1579
Qy 1605 DTDAAVQQL-NEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLG 1663
Db 1580 ADPAQQAQAEELNL-----IASAKTLSAGKKYDDAT-----TALA 1618
Qy 1664 GKGNDTEKLTAGTIGQ--VGVKDKNANGDLSNVWVTKQDKGSKKALLATVYNAAGQTNYL 1721
Db 1619 AA-----TQKAQTALDQTNASVDSLTGANRLQT-----AINQLAAKLPAKTKTSL 1665
Qy 1722 TNNPAEALDRINEGIRFFHVNDGNQEPVVOGRNGIDSSAGKHSVAIGFOAKADGAAV 1781
Db 1666 -----NQLQSVKAALGTDLGNQTDPTSTGKTFPTAALDLVA- 1700
Qy 1782 AIGRQAGNOSIAIGDNAQATGDSATIGTGNVNVAGKHSAGIADPSTVKADNSYSVGN 1841
Db 1701 ---QAQGTQT---DQHQAT---LAKVLDVLAELAEG-----IKAATPAEVGNA 1742
Qy 1842 NQFTDATQTDVFGVGNITVTESNVALGSN-----SAISAGTHAGTQAKKSD 1889
Db 1743 K---DAATGRTWYADIADTLTSGQASADADSKLAHLQALQSLTKVAAYEAATVKG 1799
Qy 1890 GTAGTTTAGATGVKFGAGTAVGAVSVGASGAERIQNVAGEV-----SATSTD 1941
Db 1800 GTTCTSDKGGQGPAPAPG-----DICKDKGDEGSPSGGNIPTNPATTTSTSD 1851

RESULT 12

P3P_LACLC STANDARD; PRT; 1902 AA.
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN-SK11;
RX MEDLINE-89340435; PubMed-2760036;
RA Vos P., Simons G., Slezien R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RL cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -I- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; J04962; AAA03533.1; ALT_SEQ.
DR PIR; A32634; A32634.
DR HSP; P00782; 2S8T.
DR MEROPS; S08.019; .
DR InterPro; IPR000209; .
DR InterPro; IPR001899; .
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 4.2%; Score 430.5; DB 1; Length 1902;
Best Local Similarity 20.5%; Pred. No. 1.4e-08;
Matches 452; Conservative 241; Mismatches 838; Indels 679; Gaps 95;

Qy 40 SVCTLSFARIAALAVLVIGATLSGSAYAKKDKTHIAIGEQNPPRRSGTAKADGDRATAI 99
Db 9 SILLAGTVALGALAVLPVGEIQAKAAISQTKGSLA-----NVTAFAKQ-----AA 57
Qy 100 GENANAGQGAIAIGSSNKTNGVSSLDKI-----GT 130
Db 58 TDTTAATTNQAATQALAAKGDYDNKLNKVKVQQDIYVDVIVQMSAAPASENGILRTDYSST 117
Qy 131 DATQES---TAIGDVKA-----SGDA-----SIAIGSDDLHLDDQHGNPK- 169
Db 118 AEIQOETNKVIAQAQAKAAVEQVTOQTAGESYGVYVNGFSTKVRVVDIPKLQIAGVKT 177
Qy 170 -----HPKCTLINDLINGHAVL-----KEIRSSKNDVKY 199
Db 178 VTLAKVYPTDAKANSMANVQAVMSYKKGEGVSVVSDIGDPTHKDMRLSDDKDKVL 237
Qy 200 RRTTASGHASTAVG-----AMSYAQG-----HFSNAGFTRATAKS 234
Db 238 TKSDVEKFTDTPVKGHRYFNSKVPYGFNVADNNDITDDKVDQHGVMHVGAIIGANGTGD 297
Qy 235 AYSLAVGLAATAEQSTIAIGSDATSSSLGAIAGAGTGAOLQGSIALGQGSVVTQSDNN 294
Db 298 PAKSVGVGAPQAQLLAKMVKFNSDTSKTS-----ATVVSATIEDSAKIGADV- 348
Qy 295 SRPAPTNPQALDPKFOATNTK-----AGPLSIGNSIKRKLIINVGAGVKNKTDVAVNAQL 350
Db 349 SLGNSNGQITLEDPELAQVQANESGTAIVISAGNSGTS-----SATEGVNK-DYIGLQDN 404
Qy 351 EAVYKWKAKERRITFGDDNDSTDVKIGLDNTLTI----- 383
Db 405 EMVSGPGTSRCATTVASAENTDV---ITQAVTITDGTGLQLGPETIQLSSDHTGTSFQDK 461
Qy 384 -----KGAETNALTD-----NNGVYKKEADNS-GLKVKLAKTLN--NLTEVNT-- 424
Db 462 KFYIVKDASGNLSKALADYTADAKGTAIVKRGESFDDKQYAAAGAAAGLIIVTDG 521
Qy 425 -----TTLNATTTVKVGSSSTTAELLSDSLTF-----TOPNTGSSQSTS 463
Db 522 TATPWTISALITTTPTTFLGSSSVTGKLVWVTAHPDDSLGKVTILMLPNQKYTEDKMSD 581

QY 464 KTVYG-VNGVKFTNAETAAIGTRITRDKIGFAROGDVDEKQAPYLDKKQKLVGSVAI 522
 Db 582 FTSYGPVSNLSE--KPDITAPCGNIWSTQNNNGYTNMSGT-SMASPFITAGSOALLKQALN 638
 QY 523 TIDNGIDAGNKISINLAKGSSANDAVTTEOLKAAP-----TLNAGAGISVTP 570
 Db 639 KNNPPYAYKOL-----KGTALTDFLKVEMTAQPIINDINYNVIVSPRQAGL--VD 692
 QY 571 TEISVDAKSNVTAPTNYTNGVKTTEINSOGTSDFSVKSGTGNNSLVTAELHASYLNEVN 630
 Db 693 VKAAIDALEKNPSTVVAENGYPAVELKDFSTDK-----TEKLTFTN 734
 QY 631 RTADSLAQSTVEEEDDDANAITVAKDTTKNAGAVSILKLGK-----NLGTVAATKDG 685
 Db 735 RTT-----HELTQYMDSNFTDNVYTSA--TDPNSGVLYDKKIDGAAKAGSNITVPAGTA 789
 QY 686 TVTFGLSODSGLTIGKSTLNDGLTVKDTNEQIQVGANGIKFTNYNGSNPGTGIANATARI 745
 Db 790 QIETLS-----LPKSFQOQFVEG-----ELNFKGS-----DGSRL 821
 QY 746 TRDKIGFAG--SPGAVDTNKPYLDDQKLOVGNVKTINTGINAGGKAITCLSPTLPSIADQ 803
 Db 822 NLPYMGFGWINDGKI-----VD-----SLNGITYSPAGNFG-----TVPLLNK 862
 QY 804 SSRNIEGNTIQDKKNSA-----ASINDILMTGFNLKNNNNPIDFVSTYDI 850
 Db 863 NTGTQYGGWTDADGKNTVDDQAIASFSSDKNLYNDISMKYLLRNISN-----VOVDI 917
 QY 851 VDFANGNATTAT-----VHTDANKTSKVYDVNVDDTTIHLTGDDNKKLGKVTTKLN 904
 Db 918 LDGOGKVTLLSSSTNRKTKYNAHSQYIYYPAPAMDGTYY-----DQDGNK 967
 QY 905 KTSANGTATNFVNSDEDAALVNAKDIAENLTAKEIHTTKGTADTALQFTVTKVDE 964
 Db 968 -----TADGGSYTYRISGVE 983
 QY 965 NNNAD-----DANAITYGQNNANQVNTLTGKENG-----LNKTKNG----- 1004
 Db 984 GGDKROQFVDFPKLDSKAPTVRHV-----ALSAPTENGKTQYILTAEAKDDLGLDAT 1036
 QY 1005 -TVTFGINNTSGLKA--GKSTLNDGGLSTKNPTGSEQIOVGADGVKFAKVNNGVVGAG 1060
 Db 1037 KSVKTEINEVNTLADFTDAGTTADGTYKIETPLSDQAALGNGDSAL----- 1087
 QY 1061 IDGTRITRDEIGFTGTGSDKSKPHLSKDGINAGG--KKITNIOGSEIAQNSHDAVT-- 1117
 Db 1088 -----YLTDNASNATDQASVQKPGSTFDLIVNGGGTPDKISSTTGT-----YEANTQ 1136
 QY 1118 GKKIYDLKTELENKISSAKTAQNSLHEFSVA--DEQGNFTVSNPYSSYDTSKTSVITF 1176
 Db 1137 GGGTYTFSGCTYPAAYDGYTDAQGGKHLNLTYYDAATNSFTASMPVTNADYAAQVDLYAD 1196
 QY 1177 ACENG1-----TPKVKGV-----VRVGIDQTKGLT-----TPKLTVGNNGKG 1215
 Db 1197 KAHTQLLKHFDKVRMLAPFTFDLKFNGNSDQTSKATIKVGTVSADTKTVNNGHTVA-- 1254
 QY 1216 IVIDSONQNTITGLS-----NTLANVTNDKGSVRTTEQGNIIKDEKTPRAASIVDLSAGF 1272
 Db 1255 -ALDAQHFSVDVPVNYGDNITKVTATDKDGNNTTTEQTKITSSYDP-----DMLKKS 1306
 QY 1273 NLOGNCEANDVFSYDT--VNEADG--NATTAKVYDDTSKTSKVYDVNVNDDTTIEVKDK 1329
 Db 1307 -----TFDQGVKFGTNKFNATSAKF--YDPKTGATITGKVKHPTTLQVQDGK 1352
 QY 1330 KLGKVK-----TTTTSTGTGANKFAL-----SNQATGDALVKASDIYVAHLNLTSGDIQT 1378
 Db 1353 QIPKDDLTFSFTLDLGLQKPFVGVVGDITQNTKTFQALSFLDAVA--PTLSLDSST 1410
 QY 1379 AKGASQANSAGVADGKNKYVDSTDNKYQAKNDGTVDKTKKVAQKLVAAQATPDGT 1438
 Db 1411 -----DAPVYTNPNFQITGATDINAQYLS----- 1435
 QY 1439 LAQMNKSVINK--EQVNDANKKG-----INEDNAFVKG-----LEKAASDNKTKNAAVTVGDL 1491

Db 1436 -LSINGSSVASQVEDININSKPGHMAIDQPKLLEGNVLTVAVTDSEDN---TTTKNI 1491
 QY 1492 NAVAQPLTFACDTGTAKKLGTELTIKGQDITDKNTLNNICVAGTGDFTVKLAKDLT 1551
 Db 1492 TVYEFKTLAAPTVPIS-----TTEPAQIVT--LTAN-----AATGETVQVSAD-- 1535
 QY 1552 NLNSVNAAGTKIDD---KGVSVFVDSGQAKANTPVLNSANGLDGGKVISNVGKTKDTDA 1608
 Db 1536 -----GKTYQDVPAAGVT--ITANGTEKFKSTDLYGNESPAVIDVVTNI---KADDP 1583
 QY 1609 ANVOOL-NEVRNLLGLGNAGNDNADGNQVNIADIKDPNSGSSNRTVIKAGTVLGGKN 1667
 Db 1584 AOLQAARQELTNL-----IASAKTILSASGKYDDATT---TALAAA-- 1620
 QY 1668 NDTFKLATGGIQ--VGVDKDGNGNDLSNVVVKTKQDGSKKALLATYNAAGOTNYLTNNP 1725
 Db 1621 --TQKAOTALDQTNASVDSLGTANRDLQ-----AINQLAAKLPADKKTSL-- 1665
 QY 1726 AEADTRINEOGIRFFHYNDGNQEPVVGRRNGIDSSASGKHSVAIGFOAKADGEAAVAIGR 1785
 Db 1666 -----NQLSVKDALGTLGNQDTPSTGKTTFAALDLVA----- 1700
 QY 1786 QTOAGNOSIAIGDNAQATGDQSIAGTGNVNVAGKHSIGAIDPSTVKADNSYSVGNNNQFT 1845
 Db 1701 QAQAGTQI---DDQLQAT---LAKILDEVLAKEG-----IKAATPAEVGNK-- 1743
 QY 1846 DATQDTDFGVGNNTITVESNSVALGNSN-----SAISAGTHAGTQAKSKDGTAG 1893
 Db 1744 DAATGKTWYADIADTLTSGQASADASDKLAHLQALQSLKTKVAAVAANAATVKGCGDTTG 1803
 QY 1894 TTTAGATGTVKGFAGTAVGAVSVG---ASGAERRIIONVAAGEVSATST 1940
 Db 1804 TSDKGGGOGTAPAPGDTGKDKGDEGSPSSGG-----NIPKPTTTTST 1848
 RESULT 13
 HLVA_SERMA STANDARD; PRG: 1608 AA.
 ID HLVA_SERMA
 AC P15320;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE HEMOLYSIN PRECURSOR.
 GN SHLA.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 OX NCBI_TaxID=615;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
 RC STRAIN=SN8;
 RX MEDLINE=88257037; PubMed=3290200;
 RA Poole K., Schiebel E., Braun V.;
 RT "Molecular characterization of the hemolysin determinant of Serratia marcescens";
 RL J. Bacteriol. 170:3177-3188(1988).
 CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.
 CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHLB FUNCTION.
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
 CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

Thu Sep 13 14:18:59 2001

```
RESULT 14
YDEK_ECOLI
ID YDEK_ECOLI STANDARD; PRT; 1325 AA.
AC P32051; P76140; P77168;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 136.5 KDA LIPOPROTEIN IN HIPB-UXAB INTERGENIC REGION
DE PRECURSOR (ORF).
GN YDEK OR ORF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 595-1325 FROM N.A.
RX MEDLINE=94100243; PubMed=8274505;
RA Cartwright P.J., Tims M.W., Lithgow T., Hoeg P.B., Hoogenraad N.J.;
RT "An Escherichia coli gene showing a potential ancestral relationship
to the genes for the mitochondrial import site proteins ISP42 and
MOM38.";
RL Biochim. Biophys. Acta 1153:345-347(1993).
CC -!- SURCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
ANCHOR (POTENTIAL)
CC -!- SIMILARITY: TO E-COLI YFAL.
CC -!- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
ISP42 AND MOM38.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 653.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000248; AAC74583.1;
CC EMBL: D90793; BAA15190.1; ALT_INIT.
CC EMBL: D90794; BAA15197.1; ALT_INIT.
CC EMBL: X73295; CAA51730.1; ALT_FRAME.
CC PIR: S34315; S34315.
CC EcoGene: EGI1780; ydek.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1325 HYPOTHETICAL LIPOPROTEIN YDEK.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
FT CONFLICT 884 884 N -> K (IN REF. 3).
```

```
FT CONFLICT 1317 1317 M -> S (IN REF. 3).
SQ SEQUENCE 1325 AA; 136514 MW; 26A3A066FA19AD7D CRC64;

Query Match 3.6%; Score 375; DB 1; Length 1325;
Best Local Similarity 23.7%; Pred. No. 9e-07; Indels 360; Gaps 83;
Matches 346; Conservative 175; Mismatches 578;

QY 658 DTTNAGAVSTLKLKGNGLVATKDKGTWTFG--LSQDSGLTICKS--TLNNDGLTVKVDTN 715
DB 21 ELTRRAGKTSIVNLRKSSGLTT---KFSRLTLGLVLLSASGASLEVDNDQITNIDTD 77
QY 716 ---EQIQVANGAIKFTNVNNGPOTGANTARITRDKIGF--AGSDGAVD----- 760
DB 78 VAYDAYLVGWYGTGVLNLAG---GNASLTITITTSVIGANEDSEGTNVNLGGTWRLYDS 133
QY 761 -TNKPYLDQDKLOVGNVKITNTGINAGG---KAITGLSPTLPSIAQSSRNIEL----- 810
DB 134 GNNARPLNVGSGTGLNKKQGHVDGGLYLSGSTGGVGVNVEGEDSVLTTELFEGS 193
QY 811 --GNTIQDKDSN--AASINDIL--NTGPN-----LKNNNPIDPVSYDIVDF 853
DB 194 YGTGSLNITDKGYVTSSIVAILGYQAGSNGQVWVEKGEWLIKNDSSIEF---QIGNQ 249
QY 854 ANGNAT-----TATVTHDTANKTSKVVDVNVDDTTIHLTGTDNKKLGVKTKLNKT 906
DB 250 GTGEATIREGGLVTAENTIIIGNATGICITLVNQDDSVI-----TVRRLYN 295
QY 907 SANGTATNFNVN---SSDEDALVNAKDAENLANTLAKEIH--TTKGATADTALQTFVVKV 962
DB 296 GYFGNGTVNISNGLINKEYSLVGVQDGHGVNVVTKGHNFLGTGEAFRIYIGDAG 355
QY 963 DENNNAD-----DANAIVGQKNANNQVNTLTAKENGINKTKDKNGTVTGINTSLK 1017
DB 356 DGEINVSSEKGYDGLIITAGMKET-----GTGNITVAK--DKNSVIT--NLGTNLGYD 403
QY 1018 A-GKSTPLNDGGISIKNPTSEIOIQVGADGV-----KFAKVNNN-----GVWGAG--- 1060
DB 404 GHGEMNISNOGLVSN--CGSSILGYGETGVGNVSITTTGGMVEVKNVYTTICVAGVGNLN 461
QY 1061 IDGTTTRITRDEITGFTGTN-----GSLDKSPHLSKD--GINAG--GKKITNIOSGEIAQNS 1112
DB 462 ISDGRKFVSONITFLDKASGIGITLMDATSFDTGVINGVNFSGSIVNVNNGATLNST 521
QY 1113 HDVAVTGCK-----IYDLKTELENKISSTAKTQNSLHEFSVADEQGNFTVSNPSSYDT 1167
DB 522 GYGFIGNASGKGIWNISDLSNLTSTNAQ--LLQGVGLGTGELNITTTGGIVKARDT 579
QY 1168 SKTSDVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVNNGKGVIVSDSONQNTI 1227
DB 580 -----QIALNDK--SKGDVRVD--GQNSLLETFNMYVGT-----SGTGTL 615
QY 1228 TGLSNTLVNTD-----KGSVRTTEGCIINIKDEDKTRAASIVDVLVSAGFNLCQNGEA 1280
DB 616 TLTNNGTLNVGEVGVLPVPAVGLNLG--AAHGAADAGFITNATKVEFGL---GEG 671
QY 1281 VDFVSTVDTVFADFAGNATTAKTVDYDTSKTSKVYVDV-----NVDDT-----TIEVKDKKL 1331
DB 672 V-FVFNH--TNNSDAGYQVDMITIGDD--KDGKVTHDAGHTVFVFNAGTYSCKTL--VNDGLL 726
QY 1332 GVKTTTTLTS--TGTGANKFALSQNTGDALVKAADIYVAH--LNTLSGDIQTAKASQANNS 1388
DB 727 TIASHTADGVGTGMSSEVTIANPGTLDILASTNSAGDYTLTNALKGDLNRVQLSSDKM 786
QY 1389 AGYVDADGNKV-----IYDSTDNKYQAOKNDGTVDKTEKAVD-----KLVAQAQPDGT 1438
DB 787 FGFTHTAGTEFAGVAGLQKDSI---FTLERDNTAALTHAMLQSDSENTTSVKVGEOSIG 842
QY 1439 LAQMNKSVINKEQVNDANKQGINEDNAPVKGLEKAAQSN--KTKNAAYT--VGDL---- 1491
DB 843 LA-MNGGTIIIFDTPAATLAEGYISVDTLVVG-----AGDTWKGRNYQVNGTGDVLIDV 897
QY 1492 -----NAVAQTPLT----FAGDTGTAKKL-----GETLTIKGGQTD-----TNK 1527
```


Db 898 PKWDPMPANNPLTTLLNLEHDSHVQVLVKAQTVIGSGGSLTLRLDQDEVEADKTLH 957
QY 1528 LIDNNIGVAGTDGFTVKLAKD-----LTNLSVNA-GETKIDDKGVSVFSDSSGQAKANT 1581
Db 958 IAQGTGVAEGDYGFRLLTAPGNGLVYGLKALNHHGGOKLT-----LAHGGAYGAT 1011
QY 1582 PVLISANGLDLGGKVISNKGKTKDTPAANVQOLNEVRNLLGLGNAGNADGNQVNIADI 1641
Db 1012 ADMSAK---IGG-----EGDLA---INTVRQ-VLSLNGQNDYQGATVYQMGTL 1052
QY 1642 KKDPSGSSSNRTV-IAKGTVLGGKGNDEKLTATGQVGV-----VDKDGAN 1689
Db 1053 RTDADGALNTRNLNSAIVDLNCTQVTFETG--QMGSTVLKFGKALTVNKGGISQ 1110
QY 1690 GDLS-----NWVYKTKQDKSKKALLATYNAAGQTNLYTN-NPAEADIRINEQGI-RFFHV 1742
Db 1111 GELTGGNLLNV---TGTLAIEGLNARYNA-----LTSISPAEVSNDLTQGLGRGNIA 1161
QY 1743 NDCNQEPVQGRN---GIDSSASGKHSVAIGFOAKADGEAAVAIGRQTQAGNQSIAIGDN 1799
Db 1162 NDG-----LTLKVTGTGELNLSISGKIVSA--TARTDVELD-----GDN 1199
QY 1800 AQATGQSTAGTGNVNVACKHSGAIGDPSTVKADNSYSVGNNOFTDATOT----- 1850
Db 1200 SRVFG--QFNIDTGSALSVEQKLNLDASVI-----NNGLLTISTERSWAMTHSIS 1248
QY 1851 ---DVEGVGNIRVTESNS-----VALGSNSAIS-AGTHAGTQAKKSDGTAG 1893
Db 1249 GSGDVKTKIGTILTLNDSAAVYOGTTDIVGGEIATGSDSAINNASQHINH--NSGVMSG 1306
QY 1894 TTTTATGATGVKGFACQTA 1912
Db 1307 NVTTAGDMNMPG--GGTA 1323

RESULT 15

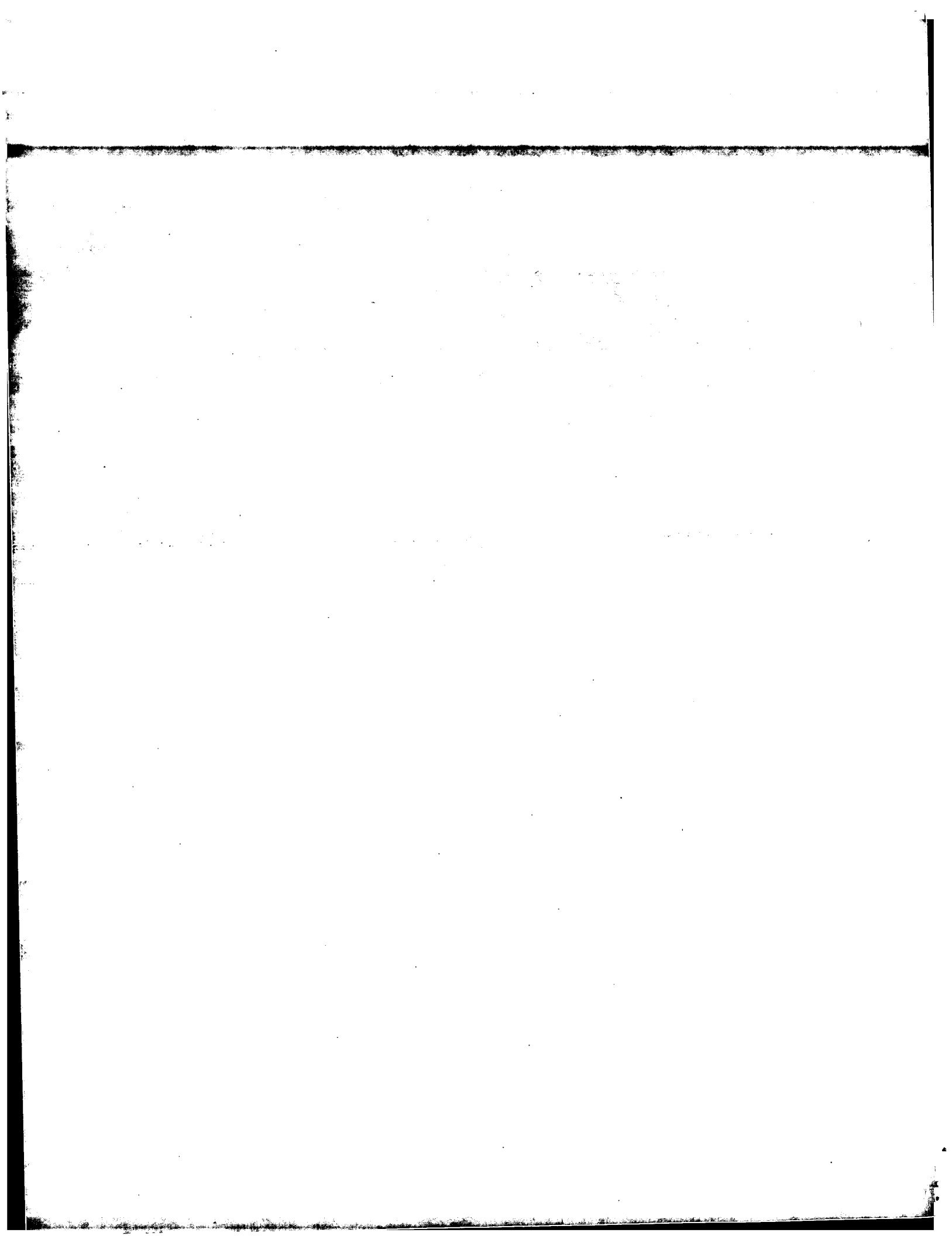
WAPA_BACSU STANDARD; PRT: 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE WALL-ASSOCIATED PROTEIN PRECURSOR.
GN WAPA OR NI7G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 238 kDa precursor two-domain ligand-binding
RT protein.";
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSCIAL;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci.";
RL Microbiology 141:337-343(1995).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE

CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L05634; AAA22883.1; -;
CC EMBL: D31856; BAA06656.1; -;
CC EMBL: D29985; BAA06260.1; -;
CC EMBL: D83026; BAA11683.1; -;
CC EMBL: 299124; CAB15959.1; -;
CC PIR: S32920; S32920.
CC Subtilisin; HG10797; wapa.
CC Pfam: PF02018; CBD_6; 1.
CC Cell wall; Repeat; Signal.
CC SIGNAL 1 28
CC OR 32 (POTENTIAL).
CC CHAIN 29 2334
CC DOMAIN 504 869
CC REPEAT 504 605
CC REPEAT 636 736
CC REPEAT 769 869
CC DOMAIN 1021 2139
CC 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
CC X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
CC 2-1.
CC REPEAT 1021 1040
CC REPEAT 1042 1061
CC REPEAT 1063 1082
CC REPEAT 1083 1102
CC REPEAT 1109 1128
CC REPEAT 1129 1148
CC REPEAT 1150 1169
CC REPEAT 1174 1193
CC REPEAT 1199 1218
CC REPEAT 1219 1238
CC REPEAT 1249 1268
CC REPEAT 1269 1288
CC REPEAT 1289 1308
CC REPEAT 1309 1328
CC REPEAT 1329 1348
CC REPEAT 1349 1368
CC REPEAT 1369 1388
CC REPEAT 1389 1408
CC REPEAT 1409 1428
CC REPEAT 1429 1448
CC REPEAT 1449 1468
CC REPEAT 1469 1488
CC REPEAT 1489 1508
CC REPEAT 1509 1528
CC REPEAT 1529 1548
CC REPEAT 1549 1568
CC REPEAT 1569 1588
CC REPEAT 1589 1608
CC REPEAT 1609 1628
CC REPEAT 1629 1648
CC REPEAT 1649 1668
CC REPEAT 1669 1688
CC REPEAT 1689 1708
CC REPEAT 1709 1728
CC REPEAT 1729 1748
CC REPEAT 1749 1768
CC REPEAT 1769 1788
CC REPEAT 1789 1808
CC REPEAT 1809 1828
CC REPEAT 1829 1848
CC REPEAT 1849 1868
CC REPEAT 1869 1888
CC REPEAT 1889 1908
CC REPEAT 1909 1928
CC REPEAT 1929 1948
CC REPEAT 1949 1968
CC REPEAT 1969 1988
CC REPEAT 1989 2008
CC REPEAT 2009 2028
CC REPEAT 2029 2048
CC REPEAT 2049 2068
CC REPEAT 2069 2088
CC REPEAT 2089 2108
CC REPEAT 2109 2128
CC REPEAT 2129 2148
CC REPEAT 2149 2168
CC REPEAT 2169 2188
CC REPEAT 2189 2208
CC REPEAT 2209 2228
CC REPEAT 2229 2248
CC REPEAT 2249 2268
CC REPEAT 2269 2288
CC REPEAT 2289 2308
CC REPEAT 2309 2328
CC REPEAT 2329 2348
CC REPEAT 2349 2368
CC REPEAT 2369 2388
CC REPEAT 2389 2408
CC REPEAT 2409 2428
CC REPEAT 2429 2448
CC REPEAT 2449 2468
CC REPEAT 2469 2488
CC REPEAT 2489 2508
CC REPEAT 2509 2528
CC REPEAT 2529 2548
CC REPEAT 2549 2568
CC REPEAT 2569 2588
CC REPEAT 2589 2608
CC REPEAT 2609 2628
CC REPEAT 2629 2648
CC REPEAT 2649 2668
CC REPEAT 2669 2688
CC REPEAT 2689 2708
CC REPEAT 2709 2728
CC REPEAT 2729 2748
CC REPEAT 2749 2768
CC REPEAT 2769 2788
CC REPEAT 2789 2808
CC REPEAT 2809 2828
CC REPEAT 2829 2848
CC REPEAT 2849 2868
CC REPEAT 2869 2888
CC REPEAT 2889 2908
CC REPEAT 2909 2928
CC REPEAT 2929 2948
CC REPEAT 2949 2968
CC REPEAT 2969 2988
CC REPEAT 2989 3008
CC REPEAT 3009 3028
CC REPEAT 3029 3048
CC REPEAT 3049 3068
CC REPEAT 3069 3088
CC REPEAT 3089 3108
CC REPEAT 3109 3128
CC REPEAT 3129 3148
CC REPEAT 3149 3168
CC REPEAT 3169 3188
CC REPEAT 3189 3208
CC REPEAT 3209 3228
CC REPEAT 3229 3248
CC REPEAT 3249 3268
CC REPEAT 3269 3288
CC REPEAT 3289 3308
CC REPEAT 3309 3328
CC REPEAT 3329 3348
CC REPEAT 3349 3368
CC REPEAT 3369 3388
CC REPEAT 3389 3408
CC REPEAT 3409 3428
CC REPEAT 3429 3448
CC REPEAT 3449 3468
CC REPEAT 3469 3488
CC REPEAT 3489 3508
CC REPEAT 3509 3528
CC REPEAT 3529 3548
CC REPEAT 3549 3568
CC REPEAT 3569 3588
CC REPEAT 3589 3608
CC REPEAT 3609 3628
CC REPEAT 3629 3648
CC REPEAT 3649 3668
CC REPEAT 3669 3688
CC REPEAT 3689 3708
CC REPEAT 3709 3728
CC REPEAT 3729 3748
CC REPEAT 3749 3768
CC REPEAT 3769 3788
CC REPEAT 3789 3808
CC REPEAT 3809 3828
CC REPEAT 3829 3848
CC REPEAT 3849 3868
CC REPEAT 3869 3888
CC REPEAT 3889 3908
CC REPEAT 3909 3928
CC REPEAT 3929 3948
CC REPEAT 3949 3968
CC REPEAT 3969 3988
CC REPEAT 3989 4008
CC REPEAT 4009 4028
CC REPEAT 4029 4048
CC REPEAT 4049 4068
CC REPEAT 4069 4088
CC REPEAT 4089 4108
CC REPEAT 4109 4128
CC REPEAT 4129 4148
CC REPEAT 4149 4168
CC REPEAT 4169 4188
CC REPEAT 4189 4208
CC REPEAT 4209 4228
CC REPEAT 4229 4248
CC REPEAT 4249 4268
CC REPEAT 4269 4288
CC REPEAT 4289 4308
CC REPEAT 4309 4328
CC REPEAT 4329 4348
CC REPEAT 4349 4368
CC REPEAT 4369 4388
CC REPEAT 4389 4408
CC REPEAT 4409 4428
CC REPEAT 4429 4448
CC REPEAT 4449 4468
CC REPEAT 4469 4488
CC REPEAT 4489 4508
CC REPEAT 4509 4528
CC REPEAT 4529 4548
CC REPEAT 4549 4568
CC REPEAT 4569 4588
CC REPEAT 4589 4608
CC REPEAT 4609 4628
CC REPEAT 4629 4648
CC REPEAT 4649 4668
CC REPEAT 4669 4688
CC REPEAT 4689 4708
CC REPEAT 4709 4728
CC REPEAT 4729 4748
CC REPEAT 4749 4768
CC REPEAT 4769 4788
CC REPEAT 4789 4808
CC REPEAT 4809 4828
CC REPEAT 4829 4848
CC REPEAT 4849 4868
CC REPEAT 4869 4888
CC REPEAT 4889 4908
CC REPEAT 4909 4928
CC REPEAT 4929 4948
CC REPEAT 4949 4968
CC REPEAT 4969 4988
CC REPEAT 4989 5008
CC REPEAT 5009 5028
CC REPEAT 5029 5048
CC REPEAT 5049 5068
CC REPEAT 5069 5088
CC REPEAT 5089 5108
CC REPEAT 5109 5128
CC REPEAT 5129 5148
CC REPEAT 5149 5168
CC REPEAT 5169 5188
CC REPEAT 5189 5208
CC REPEAT 5209 5228
CC REPEAT 5229 5248
CC REPEAT 5249 5268
CC REPEAT 5269 5288
CC REPEAT 5289 5308
CC REPEAT 5309 5328
CC REPEAT 5329 5348
CC REPEAT 5349 5368
CC REPEAT 5369 5388
CC REPEAT 5389 5408
CC REPEAT 5409 5428
CC REPEAT 5429 5448
CC REPEAT 5449 5468
CC REPEAT 5469 5488
CC REPEAT 5489 5508
CC REPEAT 5509 5528
CC REPEAT 5529 5548
CC REPEAT 5549 5568
CC REPEAT 5569 5588
CC REPEAT 5589 5608
CC REPEAT 5609 5628
CC REPEAT 5629 5648
CC REPEAT 5649 5668
CC REPEAT 5669 5688
CC REPEAT 5689 5708
CC REPEAT 5709 5728
CC REPEAT 5729 5748
CC REPEAT 5749 5768
CC REPEAT 5769 5788
CC REPEAT 5789 5808
CC REPEAT 5809 5828
CC REPEAT 5829 5848
CC REPEAT 5849 5868
CC REPEAT 5869 5888
CC REPEAT 5889 5908
CC REPEAT 5909 5928
CC REPEAT 5929 5948
CC REPEAT 5949 5968
CC REPEAT 5969 5988
CC REPEAT 5989 6008
CC REPEAT 6009 6028
CC REPEAT 6029 6048
CC REPEAT 6049 6068
CC REPEAT 6069 6088
CC REPEAT 6089 6108
CC REPEAT 6109 6128
CC REPEAT 6129 6148
CC REPEAT 6149 6168
CC REPEAT 6169 6188
CC REPEAT 6189 6208
CC REPEAT 6209 6228
CC REPEAT 6229 6248
CC REPEAT 6249 6268
CC REPEAT 6269 6288
CC REPEAT 6289 6308
CC REPEAT 6309 6328
CC REPEAT 6329 6348
CC REPEAT 6349 6368
CC REPEAT 6369 6388
CC REPEAT 6389 6408
CC REPEAT 6409 6428
CC REPEAT 6429 6448
CC REPEAT 6449 6468
CC REPEAT 6469 6488
CC REPEAT 6489 6508
CC REPEAT 6509 6528
CC REPEAT 6529 6548
CC REPEAT 6549 6568
CC REPEAT 6569 6588
CC REPEAT 6589 6608
CC REPEAT 6609 6628
CC REPEAT 6629 6648
CC REPEAT 6649 6668
CC REPEAT 6669 6688
CC REPEAT 6689 6708
CC REPEAT 6709 6728
CC REPEAT 6729 6748
CC REPEAT 6749 6768
CC REPEAT 6769 6788
CC REPEAT 6789 6808
CC REPEAT 6809 6828
CC REPEAT 6829 6848
CC REPEAT 6849 6868
CC REPEAT 6869 6888
CC REPEAT 6889 6908
CC REPEAT 6909 6928
CC REPEAT 6929 6948
CC REPEAT 6949 6968
CC REPEAT 6969 6988
CC REPEAT 6989 7008
CC REPEAT 7009 7028
CC REPEAT 7029 7048
CC REPEAT 7049 7068
CC REPEAT 7069 7088
CC REPEAT 7089 7108
CC REPEAT 7109 7128
CC REPEAT 7129 7148
CC REPEAT 7149 7168
CC REPEAT 7169 7188
CC REPEAT 7189 7208
CC REPEAT 7209 7228
CC REPEAT 7229 7248
CC REPEAT 7249 7268
CC REPEAT 7269 7288
CC REPEAT 7289 7308
CC REPEAT 7309 7328
CC REPEAT 7329 7348
CC REPEAT 7349 7368
CC REPEAT 7369 7388
CC REPEAT 7389 7408
CC REPEAT 7409 7428
CC REPEAT 7429 7448
CC REPEAT 7449 7468
CC REPEAT 7469 7488
CC REPEAT 7489 7508
CC REPEAT 7509 7528
CC REPEAT 7529 7548
CC REPEAT 7549 7568
CC REPEAT 7569 7588
CC REPEAT 7589 7608
CC REPEAT 7609 7628
CC REPEAT 7629 7648
CC REPEAT 7649 7668
CC REPEAT 7669 7688
CC REPEAT 7689 7708
CC REPEAT 7709 7728
CC REPEAT 7729 7748
CC REPEAT 7749 7768
CC REPEAT 7769 7788
CC REPEAT 7789 7808
CC REPEAT 7809 7828
CC REPEAT 7829 7848
CC REPEAT 7849 7868
CC REPEAT 7869 7888
CC REPEAT 7889 7908
CC REPEAT 7909 7928
CC REPEAT 7929 7948
CC REPEAT 7949 7968
CC REPEAT 7969 7988
CC REPEAT 7989 8008
CC REPEAT 8009 8028
CC REPEAT 8029 8048
CC REPEAT 8049 8068
CC REPEAT 8069 8088
CC REPEAT 8089 8108
CC REPEAT 8109 8128
CC REPEAT 8129 8148
CC REPEAT 8149 8168
CC REPEAT 8169 8188
CC REPEAT 8189 8208
CC REPEAT 8209 8228
CC REPEAT 8229 8248
CC REPEAT 8249 8268
CC REPEAT 8269 8288
CC REPEAT 8289 8308
CC REPEAT 8309 8328
CC REPEAT 8329 8348
CC REPEAT 8349 8368
CC REPEAT 8369 8388
CC REPEAT 8389 8408
CC REPEAT 8409 8428
CC REPEAT 8429 8448
CC REPEAT 8449 8468
CC REPEAT 8469 8488
CC REPEAT 8489 8508
CC REPEAT 8509 8528
CC REPEAT 8529 8548
CC REPEAT 8549 8568
CC REPEAT 8569 8588
CC REPEAT 8589 8608
CC REPEAT 8609 8628
CC REPEAT 8629 8648
CC REPEAT 8649 8668
CC REPEAT 8669 8688
CC REPEAT 8689 8708
CC REPEAT 8709 8728
CC REPEAT 8729 8748
CC REPEAT 8749 8768
CC REPEAT 8769 8788
CC REPEAT 8789 8808
CC REPEAT 8809 8828
CC REPEAT 8829 8848
CC REPEAT 8849 8868
CC REPEAT 8869 8888
CC REPEAT 8889 8908
CC REPEAT 8909 8928
CC REPEAT 8929 8948
CC REPEAT 8949 8968
CC REPEAT 8969 8988
CC REPEAT 8989 9008
CC REPEAT 9009 9028
CC REPEAT 9029 9048
CC REPEAT 9049 9068
CC REPEAT 9069 9088
CC REPEAT 9089 9108
CC REPEAT 9109 9128
CC REPEAT 9129 9148
CC REPEAT 9149 9168
CC REPEAT 9169 9188
CC REPEAT 9189 9208
CC REPEAT 9209 9228
CC REPEAT 9229 9248
CC REPEAT 9249 9268
CC REPEAT 9269 9288
CC REPEAT 9289 9308
CC REPEAT 9309 9328
CC REPEAT 9329 9348
CC REPEAT 9349 9368
CC REPEAT 9369 9388
CC REPEAT 9389 9408
CC REPEAT 9409 9428
CC REPEAT 9429 9448
CC REPEAT 9449 9468
CC REPEAT 9469 9488
CC REPEAT 9489 9508
CC REPEAT 9509 9528
CC REPEAT 9529 9548
CC REPEAT 9549 9568
CC REPEAT 9569 9588
CC REPEAT 9589 9608
CC REPEAT 9609 9628
CC REPEAT 9629 9648
CC REPEAT 9649 9668
CC REPEAT 9669 9688
CC REPEAT 9689 9708
CC REPEAT 9709 9728
CC REPEAT 9729 9748
CC REPEAT 9749 9768
CC REPEAT 9769 9788
CC REPEAT 9789 9808
CC REPEAT 9809 9828
CC REPEAT 9829 9848
CC REPEAT 9849 9868
CC REPEAT 9869 9888
CC REPEAT 9889 9908
CC REPEAT 9909 9928
CC REPEAT 9929 9948
CC REPEAT 9949 9968
CC REPEAT 9969 9988
CC REPEAT 9989 10008
CC REPEAT 10009 10028
CC REPEAT 10029 10048
CC REPEAT 10049 10068
CC REPEAT 10069 10088
CC REPEAT 10089 10108
CC REPEAT 10109 10128
CC REPEAT 10129 10148
CC REPEAT 10149 10168
CC REPEAT 10169 10188
CC REPEAT 10189 10208
CC REPEAT 10209 10228
CC REPEAT 10229 10248
CC REPEAT 10249 10268
CC REPEAT 10269 10288
CC REPEAT 10289 10308
CC REPEAT 10309 10328
CC REPEAT 10329 10348
CC REPEAT 10349 10368
CC REPEAT 10369 10388
CC REPEAT 10389 10408
CC REPEAT 10409 10428
CC REPEAT 10429 10448
CC REPEAT 10449 10468
CC REPEAT 10469 10488
CC REPEAT 10489 10508
CC REPEAT 10509 10528
CC REPEAT 10529 10548
CC REPEAT 10549 10568
CC REPEAT 10569 10588
CC REPEAT 10589 10608
CC REPEAT 10609 10628
CC REPEAT 10629 10648
CC REPEAT 10649 10668
CC REPEAT 10669 10688
CC REPEAT 10689 10708
CC REPEAT 10709 10728
CC REPEAT 10729 10748
CC REPEAT 10749 10768
CC REPEAT 10769 10788
CC REPEAT 10789 10808
CC REPEAT 10809 10828
CC REPEAT 10829 10848
CC REPEAT 10849 10868
CC REPEAT 10869 10888
CC REPEAT 10889 10908
CC REPEAT 10909 10928
CC REPEAT 10929 10948
CC REPEAT 10949 10968
CC REPEAT 10969 10988
CC REPEAT 10989 11008
CC REPEAT 11009 11028
CC REPEAT 11029 11048
CC REPEAT 11049 11068
CC REPEAT 11069 11088
CC REPEAT 11089 11108
CC REPEAT 11109 11128
CC REPEAT 11129 11148
CC REPEAT 11149 11168
CC REPEAT 11169 11188
CC REPEAT 11189 11208
CC REPEAT 11209 11228
CC REPEAT 11229 11248
CC REPEAT 11249 11268
CC REPEAT 11269 11288
CC REPEAT 11289 11308
CC REPEAT 11309 11328
CC REPEAT 11329 11348
CC REPEAT 11349 11368
CC REPEAT 11369 11388
CC REPEAT 11389 11408
CC REPEAT 11409 11428
CC REPEAT 11429 11448
CC REPEAT 11449 11468
CC REPEAT 11469 11488
CC REPEAT 11489 11508
CC REPEAT 11509 11528
CC REPEAT 11529 11548
CC REPEAT 11549 11568
CC REPEAT 11569 11588
CC REPEAT 11589 11608
CC REPEAT 11609 11628
CC REPEAT 11629 11648
CC REPEAT 11649 11668
CC REPEAT 11669 11688
CC REPEAT 11689 11708
CC REPEAT 11709 11728
CC REPEAT 11729 11748
CC REPEAT 11749 11768
CC REPEAT 11769 11788
CC REPEAT 11789 11808
CC REPEAT 11809 11828
CC REPEAT 11829 11848
CC REPEAT 11849 11868
CC REPEAT 11869 11888
CC REPEAT 11889 11908
CC REPEAT 11909 11928
CC REPEAT 11929 11948
CC REPEAT 11949 11968
CC REPEAT 11969 11988
CC REPEAT 11989 12008
CC REPEAT 12009 12028
CC REPEAT 12029 12048
CC REPEAT 12049 12068
CC REPEAT 12069 12088
CC REPEAT 12089 12108
CC REPEAT 12109 12128
CC REPEAT 12129 12148
CC REPEAT 12149 12168
CC REPEAT 12169 12188
CC REPEAT 12189 12208
CC REPEAT 12209 12228
CC REPEAT 12229 12248
CC REPEAT 12249 12268
CC REPEAT 12269 12288
CC REPEAT 12289 12308
CC REPEAT 12309 12328
CC REPEAT 12329 12348
CC REPEAT 12349 12368
CC REPEAT 12369 12388
CC REPEAT 12389 12408
CC REPEAT 12409 12428
CC REPEAT 12429 12448
CC REPEAT 12449 12468
CC REPEAT 12469 12488
CC REPEAT 12489 12508
CC REPEAT 12509 12528
CC REPEAT 12529 12548
CC REPEAT 12549 12568
CC REPEAT 12569 12588
CC REPEAT 12589 12608
CC REPEAT 12609 12628
CC REPEAT 12629 12648
CC REPEAT 12649 12668
CC REPEAT 12669 12688
CC REPEAT 12689 12708
CC REPEAT 12709 12728
CC REPEAT 12729 12748
CC REPEAT 12749 12768
CC REPEAT 12769 12788
CC REPEAT 12789 12808
CC REPEAT 12809 12828
CC REPEAT 12829 12848
CC REPEAT 12849 12868
CC REPEAT 12869 12888
CC REPEAT 12889 12908
CC REPEAT 12909 12928
CC REPEAT 12929 12948
CC REPEAT 12949 12968
CC REPEAT 12969 12988
CC REPEAT 12989 13008
CC REPEAT 13009 13028
CC REPEAT 13029 13048
CC REPEAT 13049 13068
CC REPEAT 13069 13088
CC REPEAT 13089 13108
CC REPEAT 13109 13128
CC REPEAT 13129 13148
CC REPEAT 13149 13168
CC REPEAT 13169 13188
CC REPEAT 13189 13208
CC REPEAT 13209 13228
CC REPEAT 13229 13248
CC REPEAT 13249 13268
CC REPEAT 13269 13288
CC REPEAT 13289 13308
CC REPEAT 13309 13328
CC REPEAT 13329 13348
CC REPEAT 13349 13368
CC REPEAT 13369 13388
CC REPEAT 13389 13408
CC REPEAT 13409 13428
CC REPEAT 13429 13448
CC REPEAT 13449 13468
CC REPEAT 13469 13488
CC REPEAT 13489 13508
CC REPEAT 13509 13528
CC REPEAT 13529 13548
CC REPEAT 13549 13568
CC REPEAT 13569 13588
CC REPEAT 13589 13608
CC REPEAT 13609 13628
CC REPEAT 13629 13648
CC REPEAT 13649 13668
CC REPEAT 13669 13688
CC REPEAT 13689 13708
CC REPEAT 13709 13728
CC REPEAT 13729 13748
CC REPEAT 13749 13768
CC REPEAT 13769 13788
CC REPEAT 13789 13808
CC REPEAT 13809 13828
CC REPEAT 13829 13848
CC REPEAT 13849 13868
CC REPEAT 13869 13888
CC REPEAT 13889 13908
CC REPEAT 13909 13928
CC REPEAT 13929 13948
CC REPEAT 13949 13968
CC REPEAT 13969 13988
CC REPEAT 13989 14008
CC REPEAT 14009 14028
CC REPEAT 14029 14048
CC REPEAT 14049 14068
CC REPEAT 14069 14088
CC REPEAT 14089 14108
CC REPEAT 14109 14128
CC REPEAT 14129 14148
CC REPEAT 14149 14168
CC REPEAT 14169 14188

Db	1400	-----LAGNDWTKRQITFTTPANAGKAVVYNEVDHDKDCKGKRAWFEVLEKEV	1451
Qy	1197	-----QTKGLTTPKLTGVNNGKGVIDSNGON-----TITGLSNT	1233
Db	1452	SSSNPVQNSSTS---ATENMNVSGASVDSSEEGFDDVSLKAARTSASQAGSVTKQTV	1508
Qy	1234	LANTVNDK-----GSVRTTEQNIKDEDKTRAAISIVDVLISAGF---NLQNG	1278
Db	1509	LGOSANDKPVYLTGCMKASSVKFTDEKDYSLQANVTYADGSTGLYNKAKPPSQEMNR	1568
Qy	1279	EAVDEFST-----YDVNPAFCGNATTAKTWYDD-----TSKTSK	1312
Db	1569	AAVVPKTPINKVDISILFQKSATGTVMFDDIRLIEGSLLT-KSYDSNGNYVTKED	1627
Qy	1313	VYVDNVND-----DFTIEVKDKLGKVT---TLT-----STGTG-----ANKF	1348
Db	1628	LGATYSTDYDETKKTSSETDAK-GEKTTYTYDOADQLTNMTLSNCTSLHSYDKEGNEV	1685
Qy	1349	ALSNOATGD-----ALVKASDIVAHLNTLSGOIQTAQKASQANNSAGYVDADGN	1397
Db	1686	SKTIRAGADQTYKFEYDVNMGKLVKTDPUG--NVLASEYDA-----NSNLTKTISPNCN	1737
Qy	1398	KV--IYDSTD--NKYYQA--KNDGTVDKTKVAKDKLVAQAQTPDCTLAQMNKSVINK	1450
Db	1738	EVLSYDGTDRVKSKSYNGTEKTIIFYDKNG-----NETSVVVK	1776
Qy	1451	EQVNDANKKQGINEDNAPVKGLEKAAADNKTNAAVTVGDLNAVAQTPUTFAGDTGTTAK	1510
Db	1777	EQ--NTTKRTFEDKNKRLTDLTORGSS-----QT-WTYPSD-----SD	1811
Qy	1511	KLGETLTIKGGQDTHKLTDDNNIGVAGTGTGFTVKLAKDLTLNLSVYNAGTIDDKGVSF	1570
Db	1812	KLKTFWIHGDKGTQOFTYKLDQMLE-----MKDSTSYSD-----	1850
Qy	1571	VDSGQAKATPVLSANGLDGKVISNVGKGTQDAAANVOQLNEVRNLLGLGNAGN	1630
Db	1851	YDENGWQT---FITNG---GGTSFS-----YDERNLVSSLHIGDKN	1887
Qy	1631	ADGNQVNIADIKDPNSSSSNRTVIK---AGTVL---GGKGNNDTEKLTGGIQGVGDKD	1685
Db	1888	--GGDILTESYEYDAN---GNRTIINSSAGKVOYEGKLNQLVKTHEDGTVIEWTYD	1941
Qy	1686	GNAGDLNVWVKQDKGSKALLATYNAAGQTNLTNNPAAIDRINEOGI-----	1737
Db	1942	GFGN---RKTVTTIKDGSSKTVNASFINMQ-----LTKVNDSEISYDKMNGR	1986
Qy	1738	---RFFHVD---GNOEPVW---QGRNGIDSSASGK-----HSVAI	1769
Db	1987	TSDGRTYTWDAEDNLATVTKKEDKPPATYKDEKGNRIQKTVNGKVTNYFYDGDLSN	2046
Qy	1770	GFOAKADGEAAVAIGRQTAQGNQSTIAIGDNAQ-----ATGD-QSIAIGT	1812
Db	2047	LYETADNNV-----TKSYTYGDSGOLLISYTEGKKFYFHYHNAHGDIIAISDT	2095
Qy	1813	GNVAGKHSGATGDPSTVKA-----DNSYSYGNNNQFDTAT-----OTDVF--	1853
Db	2096	GKTVAKYQDAMGNPTKTEASDEVKDNRYRYA-GYOYDEETGLLYLMARYEPNCGVLS	2154
Qy	1854	---GVCNNITVTESVALGNS-----AISAG---THAGTQAKSDGTAGT	1894
Db	2155	LDPDGSGLDQNGYAYGNNPNVNNVDPDGHVWLVVYNAGFAAYDGYKAYKSG-----	2209
Qy	1895	TTTAGATGTVKGFAGGTAVGVSAGASERRIQNVAAGEVSATSTDAVNSOLYKATQS	1954
Db	2210	-----KMKGAWAAA-----SNFPGKIFK-----GASRAKFTTK	2241
Qy	1955	IANATNELDRIHON-----ENKANAGISSAMAMASMPQA---YIPGRSMVTGG	2000
Db	2242	AVKITGHTRHGLNQSLCRNGRGVNLRAKINA-VRSPKKVIKQPNCATKYVGKATV----	2297
Qy	2001	IATHNGQAVAVGLSKLSDNGQWVFKINGSAD	2032
Db	2298	VLNKRGKVTIAYGSSR-AKSGHVFHTHGKN	2328

Db	409	NAKYTN-----TKPASKNIGKADVHKQWASVDVTAAYKSWNSGGA-----NYGFK	455
Qy	401	KEADNSGLKVLAKTLNLTTEVNTTUNATTVKVGSSSTTAELLSDSLTFPTQNTGSQ	460
Db	456	LHTNGG-KEYWKKLISSANSANKPYIEVYVTPKGNTPTIKAYHNGDS-----TGXF	507
Qy	461	STS-KTVYGVNGVK-FTNNAETTAAGTTRIT-----RDKIGF-AR	498
Db	508	DISMKKVEGAKGYKWIYNGKEYQAIAGSNVTWSWTKGKKLWPTSAELASRKRYKLHLDDGK	567
Qy	499	DG-----DVEKQAP--	508
Db	568	DGAELALDPSVPYKNSGSGYATSKNYNIGVSAIFDQEGEGAMSAKAPVPIPVNGKAQAPSA	627
Qy	509	-----YLDKKOLKVGVA-----ITIDNG-----IDAGKKISNLAKSSANDAVTI	550
Db	628	KYNNGNATGYFDLSWKAIVSGATGYKVQVFNKGKFETLDLGN-QTSWTTKGKKI--WPTS	684
Qy	551	EQLKAAPTLN---AGAGISVTPTEISVDKSGNVTAPTY-----NIGVKTEL	596
Db	685	AETKAGYALHLKDGSGAELPINP-----GPTYKNAGGCGAKRNYFKIIAY	731
Qy	597	NSDG-----TSDFSVKSGTNSLSVTAHLASYLN-----EVN	630
Db	732	NKDGEAISPAAATPALPIARPKNVTGYLYTN---TKSSQTYGVNLIWEKYQNAKGYKN	788
Qy	631	RTADSALQSTVKKEEDDDANAITAKDTKN-----AGAVSILKLGKNGLTV	679
Db	789	IYNGKEYQSFVDGAD-----HWTQKNIWPTSEEEKAGSYK-LHTDGGKGELA	837
Qy	680	AT-----KKDGTVFLQSQSLTIGKSTLN---NDGLIVKOTNEQIQV--	720
Db	838	LDPSPVNNANGYKGGKNSYFLVAYDANGETIPTAPNPTFHEGAEFLGTEETWSIID	897
Qy	721	-----GANGIKFTN-----VNGSNPPTGIANT--ARITDKI-----GFAGSDGAV--	759
Db	898	IPSGQLGATGNVIVNEEDLSIDRGPGLSLRTYNSLSDSBHLFGQGWYADAETSVIST	957
Qy	760	DTNKPYLDOOKLVGNVKTITWGINAGG--KAITGLSPILPSIADOSSRNIELGNTIOBK	817
Db	958	DGGAMVIDEATTHRETFK-----KADGTYPQPTGYLELTETADQF-----ILTKK	1003
Qy	818	DKSNAASINDLNTGNLK-----NNNPIDFVSTYD-----IVDFANGNATTA	861
Db	1004	DOTNAY-----FNKGGKLOKVVDGHNAT--VYTYNDKNQLTAITD-ASGRKLTF	1051
Qy	862	T-----VTHDTANKTSKVYDV-----NVDDTTHLTGTD-DNKKLGKVTTKLNKTA	908
Db	1052	TYDENGHWTSITGPKNKVTSYENDLLKKTVDGTDTGFTVSYDSEGRLVKYQSANSTEA	1111
Qy	909	-----NGWTAENFVNSDEDALV-----NAKDAENLTLAKEIHTT	946
Db	1112	KPVTEYOYSGHRLKLEKAINAKETVYVSYDADKTLTMTQPNRGKQVGYNEAGNPQVI	1171
Qy	947	KGTADTALQTFVKKVVDENNADDA--NAITVGOKNANNQVNTLTLKGENGLIKTDKNG	1004
Db	1172	DAAEGLKITNT--KYEGNNVVEDVDPNDVGTGKATESYQVD-----KDGNTSVK-DAYG	1224
Qy	1005	VTVFGINTSGLKAGKSTLND-----GGLSIKNPTGSEQIOVGADGVKFAKVNNGNVWG	1058
Db	1225	TETYEYNNNDVTKMKDTEGNVTDIAYDGLDAVSET--DQSGKSSAAVYDKYGNQIQSS	1282
Qy	1059	AGIDGTRITRD---EIGFTGTNGSLDKSKPHLS-----KDGINAGGKKTNIQSGETAQN	1111
Db	1283	KDLSASTNLKDGSEFAQKSGWNLASKDRRKISVIADKSGVLSGSKALEVLSOSTAGT	1342
Qy	1112	SHDVAVGGKIYDLKLELENKISSTAKT---AQNLSHEFSVADEQNNPT-VSNPYSSYDT	1167
Db	1343	DHCYSSATQVLEPNTTTLTSGKTKTDLAKSRATFNIDLRDQKQRIOWIHNEYS---	1399
Qy	1168	SKTSDVITFAGENG-----ITTKVN--KGVYRVGID-----	1196

Search completed: September 13, 2001, 12:52:48
Job time: 798 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:42:30 ; Search time 112.99 Seconds
(without alignments)
2396.924 Million cell updates/sec

Title: US-09-361-619-7
Perfect score: 10303
Sequence: 1 MNHIYKVFKNKATGTFMVA.....NGSADTOGHVCAAVGACGPHF 2047

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-unclassified.*
13: sp-vertebrate.*
14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1143	11.1	2059	Q9PD50	Q9pd50 xylella fas
2	1123	10.9	2353	P71401	P71401 haemophilus
3	1046.5	10.2	2712	Q9F3X5	Q9f3x5 pasteurella
4	778	7.6	1190	Q9PC04	Q9pc04 xylella fas
5	739	7.2	1299	Q9F3X6	Q9f3x6 pasteurella
6	678.5	6.6	1098	Q48152	Q48152 haemophilus
7	661	6.4	1107	Q9F2D8	Q9f2d8 salmonella
8	627	6.1	3705	Q9F285	Q9f285 yersinia pe
9	582	5.6	2340	Q9ZD91	Q9zd91 rickettsia
10	570.5	5.5	4919	Q9ZHL0	Q9zhl0 haemophilus
11	567.5	5.5	2021	Q52657	Q52657 rickettsia
12	555	5.4	2106	Q9XC47	Q9xc47 rickettsia
13	548	5.3	2586	Q9VTK8	Q9vtk8 drosophila
14	514.5	5.0	3029	Q53582	Q53582 synchocyst
15	514.5	5.0	2035	Q9XCJ4	Q9xcj4 salmonella
16	514	5.0	1604	Q9KK99	Q9kk99 rickettsia
17	512.5	5.0	1557	Q9RNI2	Q9rni2 haemophilus
18	509.5	4.9	5627	Q9I120	Q9i120 pseudomonas
19	506	4.9	4152	Q9ZHL3	Q9zhl3 haemophilus

20	502	4.9	2747	2	Q9L800	Q9l800 aeromonas s
21	499	4.8	1758	2	Q9JMS5	Q9jms5 escherichia
22	494.5	4.8	2232	5	P91365	P91365 caenorhabdi
23	491.5	4.8	13288	6	O18758	O18758 sus scrofa
24	485	4.7	2478	2	Q9LCH2	Q9lch2 staphylococ
25	483	4.7	2478	2	Q9RL69	Q9rl69 staphylococ
26	479	4.6	2468	2	Q9I2M3	Q9i2m3 pseudomonas
27	478.5	4.6	2349	2	P94750	P94750 escherichia
28	478.5	4.6	2383	2	P76347	P76347 escherichia
29	477.5	4.6	2514	2	Q9JY30	Q9jy30 neisseria m
30	469	4.6	1963	2	Q9XCQ3	Q9xcq3 salmonella
31	468	4.5	1615	2	Q9KKA4	Q9kka4 rickettsia
32	467	4.5	1255	2	Q9FDA0	Q9fd0 xanthomonas
33	466	4.5	1975	2	Q9K0S7	Q9k0s7 neisseria m
34	465.5	4.5	1616	2	Q9KKB8	Q9kkb8 rickettsia
35	462	4.5	2893	2	O25063	O25063 helicobacte
36	456.5	4.4	1029	2	O52708	O52708 rickettsia
37	456	4.4	3381	2	Q9KX33	Q9kx33 streptococ
38	455	4.4	1643	2	Q9F0P7	Q9f0p7 rickettsia
39	455	4.4	2147	2	Q9L950	Q9l950 pseudomonas
40	453	4.4	1615	2	Q9F0P9	Q9f0p9 rickettsia
41	452.5	4.4	1619	2	Q9KKB7	Q9kkb7 rickettsia
42	452	4.4	1328	2	Q9LAX0	Q9lax0 xanthomonas
43	452	4.4	1643	2	Q9F0P6	Q9f0p6 rickettsia
44	452	4.4	2055	2	O85472	O85472 abiotrophia
45	452	4.4	2703	2	Q9K0T0	Q9k0t0 neisseria m

ALIGNMENTS

RESULT 1
Q9PD50 ID Q9PD50 PRELIMINARY; PRT; 2059 AA.
AC Q9PD50:
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE SURFACE PROTEIN.
GN XF1529.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 "The genome sequence of the plant pathogen Xylella fastidiosa.";
 Nature 406:151-159(2000).
 DR EMBL; AE003982; AAF84338.1; --
 SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5DD24E10 CRC64;

Query Match 11.1%; Score 1143; DB 2; Length 2059;
 Best Local Similarity 24.0%; Pred. No. 9.6e-40;
 Matches 547; Conservative 342; Mismatches 766; Indels 626; Gaps 104;
 QY 86 SGTAA--KADGDRAIAIGNANAGQQAIAIGSSNKT--VNGSSLDKICTDATGOESTAIIG 141
 DB 86 SGTAAEQOASRNLTGSSLYVNSQGVNDVNLKNTYSIRMGSVITMTVA--GNAIAIG 144
 QY 142 GDVKASGDASIAIGSDDLHLHDHGNPKHPKGLTLNDLNGHVLKEIRSSKDNVDKYRR 201
 DB 145 SAQSSAADAL-----KASL-----A 159
 QY 202 TTASGHASTAVGAMSYAQHGFNAFGRATAKSAVSLAVGLAATAEGOSTIAIGSDATSS 261
 DB 160 TKASGARAIAGAKASADGVDTVALGSGATA-----GTGASSIAIGLNASAV 206
 QY 262 SGAIALGAGTAAQL--QGSIALGOGSVVQTSDNNSRPAYTPNTQALDPKFOATNNTKAGP 320
 DB 207 N-GAVAVGGALVTPDGAVALGNSVA--STGKLSGVDPKTKTSTDSAAKSTLAA 263
 QY 321 LSTG---SNSIK-RKTIYNGAGVYKNTDAVVAOLEAVKWKERRITPOGDNDSTDKIG 376
 DB 264 VSIGDVSSTNLKTRQLSGLAAGTSNTDAVNAQLKVVDEIASR----- 306
 QY 377 LDNLTIKGAETNALTDDNNGVVKENDSGLKVLAKTLNLTETVNTTLNATTVKVG 436
 DB 307 -GMNLTASGANGNAPGSSV-----DLKNTDKNLTITKAIG 342
 QY 437 SSSSTTAEALLSDSLTFQPTNTQSOSTSKTVGVGVKFTNNAP--TTAAIGTTRITRDK 493
 DB 343 S-----NDVQFNLNKDKVKTTLAVGDALLNTDG 370
 QY 494 IGFARDGDVDEKQAPYLDKKQLKVGSAIT-----IDNGIDAGNKKISNLAKGS---SAN 545
 DB 371 IALGTD-----VSLSTTGLAITDGPVAVTASGIDAGSKVISHVAGAVSETST 417
 QY 546 DAVTIEOLKAAKPTLNAGAGISVTPTEISYDAKSG-----NVTAPYINIG-VKTEL 596
 DB 418 DAVNGSQLNAVQ--VQASQPTFTGNEGAVKRSGLQGSVWISGESSTAGTYSGGNLKSVVD 475
 QY 597 NSDG-----TSDFKS--VKGSGTNNSLVTAHLASLYLNEVNETADSAQSFTVKEEDD 647
 DB 476 EAAGRHQLADSPKFGNVVINGGKISGVTAG-----TEETDAVNFSQLKSISTAV--- 527
 QY 648 DDANAITVAKDTTNAGAVSILKLGKNG-LTVATKKDGT-VTFGLSDQ---SGLTIGKS 702
 DB 528 DQGWTLTASGNGSKVASGTVDLKNTDGNLTISKSGSDNDVFNLSKDFKDMGTSGTT 587
 QY 703 TLNNDGLVAKDTNEQIQVANGAIKFTN-----VN 731
 DB 588 VVNDGVKV---GSDVALGTTGLTITDGPVAVTASGIDAGSKVISHVAGVVSSTDAVN 644
 QY 732 GSNPGTGANTARITRDK-IGFAGSDGAVDTNPKYLDQDKLVQGNKVTINFINAGGKAI 790
 DB 645 GSG-----LNAVQVQASQPTFTGNEGAV---KRSGLGSVVISG--ESSTAGMYSGG--- 691
 QY 791 TGLSPPLPDIADQSSRNIEL-----GNTI-----QDKKSNA--SI 825
 DB 692 -----NLKSVDEAAGRIHLQLADSPKFGNVVINGGKISGVTAGTEETDAVNFSQLKS 746
 QY 826 NDLNTGNFNLKNNPIDPVSYDVLDFANGNATTATVTHDTANKTSKVYV-----DVNVD 881
 DB 747 STAVDQGWTLTASGNGSKVASGTVDLKNTDGNLTISKSGSDNDVFNLSKDFKVD 803
 QY 882 DTTIHLTGT---DDNKKLIG---VKTTLKNTSANGNTATNFN-----VNS 920

DB 804 GMT---SGTIVVNDGVKVGSDVALGTTGLTITDGPVAVTASGIDAGSKVISHVAGVVSE 860
 QY 921 SBDAL---VNAKDIAEN-----LNTLAKET-----HTTKGT-----A 950
 DB 861 TSTDVAVNGSQLNAVQVQASQPTFTGNEGAVKRSGLQGSVWISGESSTAGTYSGGNLKSVV 920
 QY 951 DTAQTFYTKVVDEN-----NNADDANAITYGOK-----NANNOVNTLT 989
 DB 921 DEAGRIHLQLADSPKFGNVVINGGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTLT 980
 QY 990 LKGENGLINIKTKNGHIVTGTINTTSLKAGKSTLNDGGLSIKNPTGSEOI-----QVGA 1043
 DB 981 ASGANG--SKVASGGTVDL-----KNT--DGNLTISKSGSDNDVFNLSKDFK 1025
 QY 1044 DGKFEAK--VNNNGV-VGAGID-GTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGKK 1099
 DB 1026 DMTSGTIVVNDGVKVGSDVALGTT-----GLTIANG-----PAVTASGIDAGSKV 1072
 QY 1100 ITNIQGEIAQNSHDVATGGKIYDLKTELEKNKISSSTAKTAQNSLHEFSVADEQGNFTVS 1159
 DB 1073 ISHVAAGAVSETSTDAVNGSQLNAVQVQASQPTFTG-----NEGAVKRSGLQSVWIS 1125
 QY 1160 NPYSSTYDTSKSDVITFAGENGITTKVNGVVRVGDQTKGLTTPKL--TVGNNNKGIV 1217
 DB 1126 -----GESSTAGTYSGGN-LKSVVDEAAGRIHLQLA--DSPKFGNVVINGGK--- 1170
 QY 1218 IDSQNGQNTITGLSNTLANVTNKGVSVRTTEQGNIKDEDKTRAASIVDLSAGENLQGN 1277
 DB 1171 -----ISGVV-----AGTEETDAV---NFSQKLSISTAVDQGWTLTAS 1205
 QY 1278 GEAVDFVSTYDVTNF--ADGNATTAKVYDDTSKTSKVYVDVNVVD-----DTTIEVK 1327
 DB 1206 GANGSKVASGGTVDLKNTDGNLTISK-----SGSDNDVFNLSKDFKDMGTSGTTVVNN 1260
 QY 1328 DKILGVKTTTLTSTGTCANKFALSNQATGDLVKASDIADVAHLNLTSLDGIOTAKGASOAN 1387
 DB 1261 D---GVKVGSDVALGTTGLTIANGPAVTASGIDAGSKVISIV-----AGAVSETS 1308
 QY 1388 SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKREAVDKLVAQAQTPDGTLAQMNYKSV 1447
 DB 1309 TDVAVNGSQLNAVQVQASQPTFTG-NEGAVKRS--LQGSVVISGESSTAGTYSGGNLKSV 1365
 QY 1448 INKEQ-----VNDANKKQGT---NEDNAF-----VKGLEKASDNKTK 1482
 DB 1366 VDEAAGTIHLQLADSPKFGNVVINGGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTL 1425
 QY 1483 NAAVTVG-----DL-NAVAQTPLTAFAGTGTAKKLGTTLIK-----G 1520
 DB 1426 TASGANGSKVASGGTVDLKNTDGNLTISKSGSDNDVFNLSKDFKDKSITVGTOLDKDG 1485
 QY 1521 GQTDNKLTDNNIGVAVAGTDGP---TVKLAKDL'NLNSVNAVAGTKIDDKGVSVFVDSGQ 1576
 DB 1486 VKVSSNVLLDSNELVITSHSSTSSVKTLANGBSVVNRTVVNGDGVNIDB---VVVYNDLGL 1543
 QY 1577 AKANTPVLISANGLDLGGKVISVNGKGTGTDTAANYQOLNEVRNLLGLG-NAGNDNADNQ 1635
 DB 1544 SIVGGASLTLSGINAGSHKITNVTAGTEDDAVNFSQLKSSEAVDKGWTLTASGANGSK 1603
 QY 1636 V---NIADIKK-DPN---SGSSSNRTVI-----KAGTVLGGKGNNDTEKLIATGGIQVG 1681
 DB 1604 VSGGTVDLKNTDGNLAIKSGSDNDVFNLSKDFKDVETAG-----NTVNTDGVKVG 1658
 QY 1682 VDKDGNANG-----DLSNVVWVKQKSGKALLATYNAACQTNLTNPA---EATDRINE 1734
 DB 1659 SDVSLGAMGLFTANGPSVTASGFNAGDK---VISHVAVGMAOTDAVNVSQLKQAVQSVTV 1715
 QY 1735 QGTRFFHVNDGNEPVPVQGRNGIDSSASGKSHVIAIGFOAKADGEAAVAIGROTQA--GNOS 1793
 DB 1716 KATRYSTNDGG-----TOGNYDCDGTGSKALTAAGVGTQASGEGAAVGVSGAAASGKGS 1771
 QY 1794 IATCDNAQATCDGSIATIGT-----NVVAGKHG-----AIGDPSTVAKADNSV 1836
 DB 1772 TAIGRNAIASADGSVALGDGAKDGGRGAESYTGKYSQVQNTVGTVSVGDAKGETRSIS 1831

QY 1138 TAONSLH-----EFSVADEQGNFTVSNPYSSYDTSK-----1169
Db 1265 TONNKKHVTIDVAEAKVGDGLEKDTGKIKLKVDNTDGNLLTVDATKGASVAKGEFNA 1324
QY 1170 -TSDVITFAGENG-----ITTKVNGKVVVRVGDIDQTKGLT-----1202
Db 1325 VITDATTAGGTNANERGVVVGKSGNGATATETDKKKVATGVGVAKAINDAATFKVKNDD 1384
QY 1203 -----TPKLT 1207
Db 1385 SATIDSPDDCANDALKAGDGLTLKAGKNLKVVRDGNITFALANDLSVKSATVSDKLS 1444
QY 1208 VGNNGK-----GIVIDSQNGNT-----ITGLSNTLANVNDKGSVRTTQGNLIK 1254
Db 1445 LGTNGKNVITSDTKGLNFAKDSKTDGDDANIHLNGIASTLTDLTLSNGAT-TNLGNGIT 1503
QY 1255 DEKTRAASIVDLSAGFNLOG-----NGEAVDFVSTYDTVNFADGNATKATVYD- 1305
Db 1504 DNEKKRAASVKDVLNAGWVRGVKPASANNQVENIDFVATYDTVDFVSGDKRTTSVTVES 1563
QY 1306 -DTSKTSKVYDVNVDDTTIEVKDKLGVKTTLTSTG-----TGAN-KFALSNQAT-----1355
Db 1564 KONGKRETV-----KIGAKTSVIKDHNGKLTGKELKDANNNGVTVTET 1607
QY 1356 -----GDALVKASDIVAHLNLTLSGDIOTAKGASQANNSAGYVD-----ADGNKVIVDS 1403
Db 1608 DGKDEGNGLVTAKAVIDAVNKAAGWRYKTT-GANGQNDPATVAGSTNVTFADNGTGTAEV 1666
QY 1404 TDNKYYOAKNDG-TVDKTEKAV-----DKLVAQA-----QTPDGTLAQMVNYSKINKEO 1452
Db 1667 T-----KANDGSTVYKYNKVADVGLKLDGDKIVADTVTLTVADGKVTAPN-----1711
QY 1453 VNDANKQGINEDNAFVYKLEKAASDNKTKNAATVYVDLNAVATPLTATAGDTGTAKKL 1512
Db 1712 -NGDCK-----FVDSASGLADALNKLKSWTA-TAGKEGTGEVDPAISAGQ-----EVKA 1757
QY 1513 GETLTIKGGDTNKLTDNNIGVVGAGDGTVKLAKDLTNLSV-----NAG-----GTKI 1563
Db 1758 GDKVTFKAG-----DNLKIKQSKDFTYSLLKELKDLTSVEFKDANGGTGSESTKI 1808
QY 1564 DDGVSFVDSGQA-----KANTPVLNANGLDLGGKVISNVGKTK-----DTDAA 1609
Db 1809 TKDGLTITPANGAAGAGANTANTISVTKDISAGNKAVTVVSGLKKFGDGHFLANGTVA 1868
QY 1610 NVQQL--NEVNLGLGNAGNDN-----ADGNQVNIAD-----IKKDPNSGS-----1649
Db 1869 DFEKHYNAYKDLTNLDEKAGDNNPTVADNTAATVGDRLGLGWVISADKTTGEPNQEYNA 1928
QY 1650 ---SSNRTVIKAGT-----VLGG-----KGNNDTEKLTGCG-----IQVG 1681
Db 1929 QVRNANEVFKSGNGINVSGLTNGTRVITPELAKGEVYKSNFETVKNADGSETNLVKVG 1988
QY 1682 -----VDKDNANGDLSNVVKTOK-----DGSKKALLATYNAAGOTNYLTNN 1724
Db 1989 DMYSKEDIDPATSKPMTG---KTEKYKVENGVKVSANGSKTEVTLINKSG---YVTGN 2042
QY 1725 PAEADIRINEQGIREFHVDNGQEPVVOGRNGIDSSAGSKHSVAIGFOAK-----1774
Db 2043 ---QVADATAKSGFEL-----GLADAAEAEKAFESAADKDKLSDKAETV 2084
QY 1775 -ADGEAAVAIGROTQAGNOSTAIGDNAOTGDO-----SIAIGTGNVAGKHSAGIG 1825
Db 2085 NAHDKVRFANGLNTKV---SAATVVESTDANGDKVTTTFVKTDVDELPLTOIY---NTDANG 2138
QY 1826 DPSTVKADNSYSGVNNNOFTDQTDVFGVGNNTITVTESNVAGLSNAISAGTHACTQA 1885
Db 2139 NKIVKADCKW-----YELNADGTASKEVTLGNVANGKK-----VV 2176
QY 1886 KKSDDGTAGTTTTAGATGTGVKGAGTAVGAVSVG-----ASGAERRIONVAG 1933
Db 2177 KVTENGADKWKYTNADGAADKTKGEVSDNDKSTDEKHHVRLDPNNQNGKGVVIONVANG 2236

QY 1934 EYSATSTDAVNGSOLY---KATQSIANATNELDHRHONENKANAGISSAMAMASKPOAY 1990
Db 2237 ETSATSTDAINGSLQHYAVAKGVTNLAGQVNNLEGVKNVKGKRADAGTASALAASOLPOAT 2296
QY 1991 IPRSMVTGGIATHNGQGAVALGSLKSLSDQGVWFKINGSADTOGHVGAAGVAGGFHF 2047
Db 2297 MPCKSMVAITAGSSYQOGLAIGVSRISDNGKVIIRLSGTTTNSQGRKTGVAAGVGQW 2353
RESULT 3
Q9F3X5 PRELIMINARY; PRT: 2712 AA.
AC Q9F3X5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MAPB PROTEIN.
GN MAPB.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ277636; CAC14203.1;
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;
Query Match 10.2%; Score 1046.5; DB 2; Length 2712;
Best Local Similarity 22.9%; Pred. No. 1.5e-35;
Matches 591; Conservative 348; Mismatches 917; Indels 725; Gaps 119;
QY 7 VIFNKATGTFEMAVAEYAKSHSTGGSCATGQVGSVCTLSFARIAALAVLVIGATLSGSAY 66
Db 309 VYDGASSVYVLPKDGESSVAIGDKAVSRGE-----ASIAIGK-----NAI 349
QY 67 AOKDKTK-----HTAIGEQNPPRSGTAKA-----DGDRAIATIGENANAAGGQAI 111
Db 350 TSNDAKASQIAKNNIALGTNAQAIASDNSIALGNARTNKDSDSIAIGDAETKAAHSL 409
QY 112 ATGSSNKTVNGSSLBKIG--TDTAQESTAIAGDVKASGDASIAIGSDDLHLDDO-----164
Db 410 AVGTTSKALAEAL-ALGKLAETKTSVAMGNTSKADGNSVAVGNTSQTLSQNTIAIG 468
QY 165 HGNPKHPKGLTLNDLINGHAVLKEITRSSKDNVDKYRRTTASG-----HAST-AV 212
Db 469 SSAIANPERTISIGLNAGKQEDATGTHSQINIGENSGEGVIGQLNIGIGHAGTNNV 528
QY 213 GAMSVAQGHFSNAFGTRATAKSAYSLAVGLAA-----TAEGQSTI-----AI 254
Db 529 GKHNITAGTYAGTNLKNSEETSGANVSIGHEANKYDQLTAVOKSTVVGQAOTKAASRSTAL 588
QY 255 GSDATSSLGATAGTATRAQGLGSLGALQGGSVVTVQSDNNSRPATPTQALDPKFOATN 314
Db 589 GAETALGLDAVAVGITSKAEGDKSVAIGANS--TADSNVALGATSRVAKEGSGYLIG 646
QY 315 NTKAGPLSLG-----SNSIKRKIIINVAGVKNKTDANVNAQLEAV-----353
Db 647 KQSSLVVSVGHQDGAADQHLRLRVNVDGVDEODATVAQLKKYKTEKVTSDLOAQFNAL 706
QY 354 -----VKW-----AKERRITFGOD-----DNSTD-VKGLDNTLTIKG 386
Db 707 SHAPTASEIKYDTPVPSPTTGAENKITLAKTRISNVAPAEELDTDAVNLGVNHHVTKNK 766
QY 387 AETNALTDNNIGVYK-EADNSGLKVKL-----AKTLNNLTVEYNTTLNATTTVKYG 436
Db 767 AHYFVSNDTGINVPVGNHNDGATAKLMAIGONAARAAKERSVAIGNNTVNGEGSIGLG 826
QY 437 SSSSTTAEALLSDSLTFTQPTQNTSQSTSKTVYGVN-GVKFTNNAETAAIGTTTITRDKIG 495

Db 827 TYVKGSAE-LDDGVPRESETTVTKPSKNVXGIAIGACTTTDCGNSIAIGSLAATSDK-- 883
QY 496 FARDGDVDEKOAPYLDKKOLKVGSAITIDNGIDAGNKKISLNKAGSSANDAVTIEOLKA 555
Db 884 -----NPGANVDRAIAIGVNAVSSAFKANAIGRAVANSVKG 920
QY 556 AKPTLNAGAGISVPTPEISVDKSGNVTAPTNYIGVKT----- 593
Db 921 NAFGSQALSG-ABESSTAIGTESSEGGNA--YALGTSHAKGLNSIAFGTNQVYVSGQNSG 977
QY 594 -----TELSNDGTSOKFSVKGSGTNNSLVTAEBHLASY--LNEV----- 629
Db 978 SIGYAGELGNAKATVINCEGT---YSL---GNTNSTUTANESGIFGNSNEIKAKENARIV 1031
QY 630 -NRTADSA-----LQSFVKEEDDDANAITVAKDRTTKNAGAVSILK 670
Db 1032 GNKNTIGAIEBKPHVPGTTPAAPVNDLKDIIYVTGYDNKISSDKKLAKD-----LSGLF 1084
QY 671 LKKG-NGLTVATKKDGTIVTGLSQDSGLTIGKSTLNDGL-----TVKDTNE-QIOVG 721
Db 1085 VYGHGNIAQLPDPDSTEEFTLT-DSSVIGANNLTNTKGRNYFVLGNVNTATLNSVYLG 1143
QY 722 ANGKFTNVNNSPGTGIANAR-ITRDKIGFAGSD--GAVDTNKPYLDDOKLQVGNVKI 778
Db 1144 ADS-AYTTGNTSMNSYADMANGLNKGVTTFAGSQPVGVTVGAVGKERRVONVASGLV 1202
QY 779 TNRGNA-GKKAITGLSPTLPSIADQSSRIELGNTIQQDKSKNAASINDILATGFNL-- 835
Db 1203 TEASTDAINGSOLFALTRPLRFAGDNST---LSNP-NGKPGDVTVISRSSNOGMKVVG 1257
QY 836 -KNNNPIDFVSYDVIDFANGNATVATVTHDTANKTSKVYVVDVDDTHILTG-TDDN 893
Db 1258 GENDGNKLTITADKNIGVANGDHTLEVRKATLSNLKDATFGTGTDKTINKDGMTITN 1317
QY 894 KKLGVKTTKLNKTSAN--GNFATNFVNSSDEDAVNAKDI---AENLNTLAKEIHTK 947
Db 1318 ---GANTVSLTEGLNGMGNKINVAAGQNETDA-VNVRQLNDLKAEFGGLTGEDGGTVK 1373
QY 948 ---GTA-----DIALQFTYKVKVDENNADDA--NAITVG--OKNANNQVN-TITLKEN 994
Db 1374 QALGTAIKVTDGDDNVKTKIVTDADGSKKLEIGLENQVTLGGEAKGNPAADGKLTLANQA 1433
QY 995 GLNIKTDK-----NGTVTGTINTSG-----LKAGKSTLNDGGLSIKNPTSEQIOVG 1042
Db 1434 G-----TDKVVLDGANGTV--GLTGADGAQAQVITVKGRPTLDNAAETPRIAYGNEEVATL 1487
QY 1043 ADGVKF-----AKVNNN-GVUGA-----GIDGTTIRTRDEI----- 1072
Db 1488 NDGLKFGANAGDVHNAKINTQVDVKGATVNDVNDPDKQONIMTRVBGNTITVALAKALS 1547
QY 1073 -----GFTGTNGSLDKSKPHLSKDGINAGGKKITNIQSGEIA 1109
Db 1548 GLTSATFGDPASNPKDSVINKDGLTITQGDNTVS---LTDGLDNGNKQIKNVASGLTT 1604
QY 1110 QNS-----HDVAV--TGCKIYDLKTELENIKSISTAKTAQNSLHFSFVADSGGNTFTVSNP 1161
Db 1605 TNGTATFTSLDPAVQTGVNNGVGLKLTAINNITNGT-----NPLGGFGKLDKAGNTF----- 1654
QY 1162 YSSVYDTSKTSDDVITFAGENGITTKVNGVVRVGIDQTKGLTTPKLTVCNNKNGKIVIDSQ 1221
Db 1655 --KONLGETAQI--TGDSNVNTKV-----VD-----QONGGKALEVSL- 1688
QY 1222 NGQNTITGLSNTLANVNDKSGSVRTTEBQNI IKDEKTRAAISIVDLSAGFNLOQNGEAV 1281
Db 1689 --ANQLTGLGKPEANVPNATG---EAGKITLKDQKGTDRVVVDGSEGAISLTGQPATQ 1741
QY 1282 DFVSTYDVTNPAQGNATKAVTYD-----DTSKTSKVYVDV-NVDDTTI----- 1324
Db 1742 GAAAPTAKIRVAEGNPDLENTSDDPANPNQNKTRITRYDIAGPNGTVTVEQLATLNDGLK 1801
QY 1325 -----EVKDKKLGVK-----TTTLTSTGTCANKFALSNOATGDALYKASDIVAHLN 1370

Db 1802 FGANTGDVHDAKLNTRVDVVKGAENTWNPADACQ--IMTQISGNTIT-----VALAK 1853
QY 1371 TLSGDIOTAKGASQAN-----NSAGYVDADGNKVI-----YDSTDNKYYQA----- 1411
Db 1854 ALAG-LDSATFGNPDAGSKDGAVINKDKLTITEGDKTVKLTKEGLDNGGNOIINVDGLK 1912
QY 1412 KNDGTVDKTEV-----AKDKLVAQAQTFDGTGLAQ 1441
Db 1913 KTDGSSVALKDAEGSVLTNGVNVGDLKNAIKDVTSATNGFGFLKDKAGAEFKDGLGTAAQ 1972
QY 1442 MNVKSVINKEOVNDANKQ-----GIN-EDNAFV----- 1469
Db 1973 ITGDKNINTKVIDVPNSNDKALEISLANDITLTKNGADGVDSGLGVNGKOGASVVLNGKD 2032
QY 1470 -----KGLEKAASDNKTKNAAVT-----VGLDNAVA 1495
Db 2033 GSIGLTGPRGQDSDGSKSATISVKDGRAGVDGDKGDKTRIVYETKDATGKPVVEEVATL 2092
QY 1496 QTPLTFAAGDTG-TTAKKLGETLTIKGG-----QTDTKLTONNIG----- 1534
Db 2093 NDGMKFVNDGKEVTRKLNELTLDIKGGLDAATVADNAKVVSSNLGVKTNAEGTGLEIYMK 2152
QY 1535 -----VWAGTDG--FTVKLAKD-----LTNLNSVNAAGSTKIDDK---GVSVF 1571
Db 2153 ERPTFSLVWNGKDGEDAAVAFKADGDKDGMIAAVTD-NOGNATGLTITKDKDGNPVTFF- 2210
QY 1572 DSSQAKANTPVLSSANGLDLGGKVISNVGKGTDKDTDAANVOQLNEVRLLGLGNAGNDNA 1631
Db 2211 NNDGR-----ITNVTAGVDDKDAVNVSQLD-----GLAKATTKE 2246
QY 1632 DGNQVNIA-DIKKDPNSGSSS-----NRTVIKAG-TVLGGKGNNDTEK----- 1672
Db 2247 AGKNMTVPTVNOQ--GSTTYTVATEDNVNFTTGTGNTVM---NNDGVKVGDNVALT 2299
QY 1673 -----LATGGIQGVGVDK-DGNANGDLS-----NVWVKTKDGSKKALLATYN 1713
Db 2300 NEGLKAGDVTVTAGINAGNKKVTGVADGDISPNSTDAVNGSQLNAVKAETAEAGWHLLTAN 2359
QY 1714 AAGOTNYLTNNPABEIDRINEQG-IRFFHVNDGNOEPVYVQGRNGIDSSASGKHSVAIGFQ 1772
Db 2360 GADSSNV---KPRNTVDLNTDGNIVISKTNTADKHNT---FGLADNINVKDSVVVGP 2413
QY 1773 AK--ADGEAAVAIGRTOQA-GNQSIATGDNAQATGDSIAI-----GTGNV-----VAG 1818
Db 2414 GANGKPEGAVVINAEDGANGKDGISI---VGKDGKDAVAISGDKGVGTIGLTPAGADG 2470
QY 1819 KHSQA-IGDPSTVKADNSYSVGNNOQFTDATQTDVFGVNNITVTESNSVALGNSALS 1877
Db 2471 KNAIILGVNDVSVAGLDGNDGKOGKNSKTRIVYTPNGEEBQV-ATMNDGLVFGADKGTGH 2529
QY 1878 GTHAGTOAK-KSDGTAGTTTTAGATGTVKGFAGQTAGAVASVVGAS-----GAERRIQH-- 1929
Db 2530 KAKLGTTVKVGDKDKNIETEAVAGDTIRVR-LKDNIDVKGINVTENLTVKEGAKINMGNNV 2588
QY 1930 ---VAAGEVSATSDAVNGSOLYKATQSI---ANATNELDRIHONENKANAGISSAMAM 1983
Db 2589 IDGVADGEVATNSQAQVNGSQLHRVQOQVNNQATAINKLGDHINKVKDKDLRAGIAGATAV 2648
QY 1984 ASMPQAVIPGRSMVTGGIATHINGOGAVAGLSKLSDNGOWFKINGSADTOGHV--GA 2041
Db 2649 AFLORPNEACKSVLSGVGYSRESAATAVYARNSDNNKISIKLGGGMNSRKGVDNFGGSI 2708
QY 2042 G 2042
Db 2709 G 2709

RESULT 4

Q9PC04

PRELIMINARY; PRT; 1190 AA.

AC Q9PC04;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

[illegible]

Qy 1511 KLGETLFIKGGQDTNKLTDNNIGVWAGTDCGFTVKLAKDLTNLNSVNAAGTKIDDKGVSF 1570
Db 796 SMGLVI-----TDG----- 804
Qy 1571 VDSGQAKANTPVLISANCLDGGKVISNVGKTKDQDAANVOOLNEVRNLLGLGNAGNDN 1630
Db 805-----PVTSSGINAGSKITNVAAGTADTDAVNLSQLN----- 838
Qy 1631 ADGNQVNIADIKDPNSGSSNRTVIRAGTVLGGKGNNDTEKLT--GGIOGVVDKQDGA 1688
Db 839-----TAMACSGAKSVHYSTYDGGTQ-----GGNY 864
Qy 1689 NGDLSNVVWTKQDGSKKALLATYNAAGQTNLYLNNPAEADIRNEQOIRFFHWDGNQE 1748
Db 865 NGD----- 867
Qy 1749 PVVOGRNCIDSSAGKHSVAIGFOAKADGEAAVAIGRQTOA-GNQSTAIQDGAQATGDOS 1807
Db 868-----GATGTRSIAGVGTLASAGATAVGSAAGSAGKSTAIQGRNVAASADGS 916
Qy 1808 IAGTGT-----NVVAGKHS-----AICDPSTVKADNSYSVGNNNQFTDAT-- 1848
Db 917 VALGDGAKDGARGAESYTGKYSGLQNTVGTVSGDASKGETRTVSNVADAKEATDAVNL 976
Qy 1849 -QTVFG-----VGNNI-----TVTESNVALGNSA--ISAGTHA-----GTOAKK 1887
Db 977 RQDRVAQDANRYVDNKIESLSEGTQFKVNSL--NNSATPIAAGVDATAIGVGATASG 1033
Qy 1888 SDCTA-GTTTATAGATGVVKGFGATG-----AVSVCASCAERRIQNVAAGEVSA 1937
Db 1034 ADSAMNKASADNAV-----AIGNHSVADRANTVSVGSAGSERQVTVNVAAG---T 1083
Qy 1938 TSTDVAVNSQLYKATQSTANATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSMV 1997
Db 1084 ADTDAVNSQL--NQGLITAKQYTDGVVGSRLRDTDGVVAAAATANLPAAYIPGRGMT 1140
Qy 1998 TGGIATHNGQAVAVGLSKSDNCOWFKINGSADTOGHVGAAGVGFHF 2047
Db 1141 SVGVSSYRGQSAIAGVSSVSESRWVFKFSGANSRQVIGAGVGYQW 1190

RESULT 5

Q9F3X6 PRELIMINARY; PRT; 1299 AA.
AC Q9F3X6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MAPA PROTEIN.
GN MAPA.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277635; CAC14202.1; -
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;

Query Match 7.28; Score 739; DB 2; Length 1299;
Best Local Similarity 20.88; Pred. No. 3.9e-23;
Matches 443; Conservative 231; Mismatches 545; Indels 908; Gaps 84;

Qy 1 MNHIYKVFINKATGTFMVAEYAKSHSTGGGSCATG---VGSVCYTLTSFARIAALAVLVI 57
Db 1 MNKIYRILWNAQTOSVWVVELAKA---GKKSAGKSALVNSVSGFSFTLIAASVVL-- 54

Qy 58 GATLSGSAYAKKDKTHIAIGEONPRSGTAKADGDRAIRAIGENANAGQOIAIGSSN 117
Db 55-----GS-----GQVNAAEVTGNTGVSGDDKYCF---YNASSQSVICGDATT 93
Qy 118 KTVNGSSLDKIGTDATGOESIAIGDVKASGDASTAIGSSDDLHLLDQHGHPKPGKTLIN 177
Db 94 KT-----TDKT----- 99
Qy 178 DLINGHAHLKEIRSKONDVKYRRTTASGHASTAVGAMSYAQGHFSNAGTTRATKAYS 237
Db 100-----DN-----KPKAS 106
Qy 238 LAVGLAATAEQGSTAIGSDATSSSLGAIALGAGTRAOLOQSIALGQSVVYQSDNNS-- 295
Db 107 VVIGFATNDEGTNVAIGAKSKSKAASIAIGDNAKALDNOAIAIGQATANSWDIDISTG 166
Qy 296 RPYATPNTQALDPFQATNNTKAGPLSIGTSIKRKLINVGAGVNTKDAVNVQAQLEAVVK 355
Db 167 ROAGAEQTEV-----SAGRNIAIGDGALKR-----GKGVN----- 197
Qy 356 WAKERRITFOGDDNDVDKIGLDNTLTIKGAETNALTDNNIGVVYKEADNSGLKVKLAKT 415
Db 198-----NNIALGTSAGD-----RLAGT 213
Qy 416 LNNLTVNTTTLNATTVYKVG-SSSSTTAELLSDSLTFQPNVTGSOQSTKTVYGVNGVKF 474
Db 214 HNVLM--GTYVNADEAVRSALTAGSTKEI-----NAKETDKNKYI----- 253
Qy 475 TNNATTAAGT---TRITRDKIGFARDGDVDEKQAPYLDKKQLKV-GSVAITIDNGID 529
Db 254--EASNTVALGTRALATLAAVAIG-----QQAQAFGNOSVAVNGCTK 294
Qy 530 AGNKISNLAKGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISVDKSGNVTAPTNYI 589
Db 295 ASGO--TAIATGSAH-----ATGSSSIAIGGTVNGNDTARTLASDNFTTA--M 339
Qy 590 GVKTELNDSGTSKFSVKSGTNNLSVTAHLASYLNEVNTASALQSFTVKEEDDDD 649
Db 340 GLSAQATKSDAIAVGRNAKAAGINTVSI-----GYNAGVTKTAD-----EAQSDP 384
Qy 650 ANAIVAKDIT--KNAGAVSILKLGKNGLTVAATKDGTVTFGLSQDSGLTIGKSTLNN 706
Db 385 NKKLEPATDAVFIGNKAGYKS-----NQRMOVSLGKDSGE--GVVGTENVVTIGNSAGKN 437
Qy 707 DGLTVKTDNEQIQVGANGIKETPNVNGSNPGTGIANTRITRDKIGFAGSDG----- 757
Db 438--TKGNTNVAISSRAG---QNVGHD-----NFAALIEAGONIKGSDNIAICKHAGR 484
Qy 758 AVDTNKPYLDDQDKLOVG--NVKITNTGINAGKAITGLSPILPSIADOSSRNIEL--GNT 813
Db 485 SADPNTKLINNTISLKGESVSLKNFGIAQGNKAKT---DGLASIA--IGRNAEAVGGET 539
Qy 814 IQDKDKSNAASINDILNTGFLNKNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTSK 873
Db 540-----ANIA-----IGDSASADASGAIVLGTAKAQSLS 567
Qy 874 VVYDNNVDDTTHLTGTDNNKKLGKVTTKLNKTSANG-NTATNFNVNSDEDAL-VNAKD 931
Db 568 TV-----DGKKYGAYSAIVICTEAKATAQAAPAGKNENPKDAIAGTKA 611
Qy 932 IAENLNTLAKEIHHTTKGTADTALQFTTVKKVDENNADANAITYGQ--KNANNOVNTUT 989
Db 612 EAHVSTIALGFGAKSDTKAQAV-----SIGYSNNAKYQAIATGSEAKTTENAGSSIA 665
Qy 990 LKGENGLNIKTDKNGTVTFGINTTSLKAGKSTLNDGSLSKNPTGSEQIQVG----- 1042
Db 666 F-----GTRAKQTRASASIAIAGMGAETG-----DGGQAL---DGSDAVALGREAKAR 710
Qy 1043 ---ADGVKFAKVNNGV-VGAG-----IDGTRITRDEIGFTGTNGSLDKSKPHLSKDG 1092
Db 711 QNALAFGKAVADHKDAVALGAGAEATAAEGTNEATVNEFKYSFGAG---IKPAT--- 763
Qy 1093 INAGCK-----KITNIQSGEIAQNSHDAVTVGGKIYDLKTELENKISSTAKTAQNSLHEFSV 1148

Db 764 VSVGKKDAERTITNVAAGRDKTSTDAINGSOLY-----LALNAL----- 803
Qy 1149 ADEQGNFTVNPSSYDFSKTSDVITFAGENGITTKVNGVVRVGDQTKGLTPKLV 1208
Db 804 ----- 803
Qy 1209 GNNNGKVIDSONGONTITGLSNTLANVTNDKGSVRTTEOGNIIKDEDKTRAASIVDV 1268
Db 804 GN-----VGNLT--VTNVLGG-----DAAIKKEGDE----- 827
Qy 1269 SAGFNQNGEAVFVSTYDVFNFADGNATTAKVYDDTSKTSKVYDVNVDVDTTIEVKD 1328
Db 828 -AG----- 829
Qy 1329 KKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDITOTAKAGSOANNS 1388
Db 830 -----TLTMSNIGG-----TGKG----- 842
Qy 1389 AGYVDAGNKVYDSDTNDKYYQAKNDGTVDKTEVAKDKLVAQAQTPDGTLAQMNKSVI 1448
Db 843 -----TIHDA----- 852
Qy 1449 NKEQVNDANKKOGINEDNAPVKGLEKASDNKTKNAAVTVGDLNVAQAQPLTFAGDTGTT 1508
Db 853 NTAASKTTVEG---DNITV--TEEAAD----- 877
Qy 1509 AKKLGETLTIKGGQDTNKLTDNNIGVVAGTGFYKLAKDLTNLSNVNAGTKIDDKGV 1568
Db 878 -----GSRITYVATKDKV--KFDVSVAGGTFIDANGL 907
Qy 1569 SFVDSSGQAKANTPVLSANGILGKVLISNVKG--TKDT--DAANVQOL-----NEVRNLL 1621
Db 908 TFVDDQGTIDNTPISKTGIDAGQKVTWQNGNIAKSDKDAVNGQOLFAQEGVKNI 967
Qy 1622 GLGNAGNDNADGNQVNIADIKKDPNSSSSNRVTVIKAGTVLGGKGNNDTEKLATGGIQVG 1681
Db 968 G-----GDTTYNPETGEYAN-----TNIGGTGASTI----- 993
Qy 1682 VDKGNANGDLSNVWTKQDKSKKALLATYNAAGOTNLTNNPAAIDRINEQGRFFH 1741
Db 994 -----DEAIKAVNTAKAA----- 1007
Qy 1742 VNDGNQEPVVGNGRIDSSAKSHSVAIGFOAKADGEAAVAGROTQAGNOSIAIGNAQ 1801
Db 1008 -----KTEVQGENIVVTSAPGAN-----GNTVYTVATAKE 1038
Qy 1802 ATGDQSIAGTGNVYAGKHSAGIDPS--TVKADNSYSGN--NNQFTDATQTDVFGV-- 1855
Db 1039 VTDFDKTT-----VGSVYTDKNNDITGLSKTLGGDNFAKNGRAASEEQLNATQTNLATLLG 1095
Qy 1856 GN-----NITVTESSNALGS--NSAISAGTHAGTQAKK-----SDGTAGTTTTAGATGT 1903
Db 1096 GNAQNTNGNVAMTDIGGTCNNINDAIKASRNEVKQGNMNVVPTPTTGANGQTIYEVAATD 1155
Qy 1904 VKGFAGQTAAGVAVSGCAERIQNVAAGEYSATSTDAVNGSOLYKATQSIANA---TN 1960
Db 1156 KVAF-DEVKVGITIDAT--TKRISGIAKGDISENSTDAVNGSOLYELQOKIAKSGDNYN 1212
Qy 1961 ELDRHTHONENKANAGISSAMAMASMPQAYIPGRSMVTTGGIATHNGOGAVAVGLSKLSDN 2020
Db 1213 ILNRRINKVDKDLRAGIAGANAAAGLPQAYIPGKSNVAAAGTYKQNAIAGMSRISDN 1272
Qy 2021 GQWFKINGSADTQGHVGAAGVAGFHF 2047
Db 1273 GKVIKILKNTNSRGDFGAGYQW 1299

RESULT 6
Q48152 Q48152 PRELIMINARY; PRT: 1098 AA.
AC Q48152;
DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE ADHESIN (H1A).
GN H1A.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NONTYPEABLE STRAIN 11;
RX MEDLINE=96332658; PubMed=8730864;
RA Barenkamp S.J., St Gene J.W. III;
RT "Identification of a second family of high-molecular-weight adhesion
RT proteins expressed by non-typable Haemophilus influenzae.";
RL Mol. Microbiol. 19:1215-1223(1996).
DR EMBL: U38617; AAC43721.1;
SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;

Query Match 6.6%; Score 678.5; DB 2; Length 1098;
Best Local Similarity 23.6%; Pred. No. 1e-20;
Matches 304; Conservative 157; Mismatches 428; Indels 399; Gaps 50;

Qy 944 HTTKCTADTA---LQTFVTKVVDENNADDANAI--TVGQKNANNQVNTL-----TL 990
Db 26 HTKASATVAVAVATLALLSATVEANNPTVTKLKAYGDANFNFTNSIADAEEKOVQEA 85
Qy 991 KGENGLIKTKDKNGTFTVTEGINTTSGLKAGKSTLNDGGLSINKPTGSEIQIAGDGVKFAK 1050
Db 86 KLLNLLN---EKNASDKLLVEDNTAATVGNLRKLGWLVSSKNGTRNEKSO-----QV 137
Qy 1051 VNNNGVVGAGIDGTTTTRTDEIGFTGTNGSLDKSPHLSKDGINAGGKKITNIQSGEIAQ 1110
Db 138 -----DEVLFEG-----KGGVQVTS----- 152
Qy 1111 NSHDAVTGKIYDLKTELENKISSTAKTAQNSLH--EFSVADEQGNNTVTNPSYSDTS 1158
Db 153 -----TSNGKHITTEALAKDLG-----VKTA 174
Qy 1169 KTSVDVITFAG--ENGITTKVNGVVRVVGIDQTKLTPKLV--GNNGKGVIDS--QNG 1223
Db 175 TVSDTLTIGGGAAGAT-----TTPKVNVTSTTDLGKFAKDAAGANG 216
Qy 1224 QNTI--TGLSNTLANVTNDKGSVRTTEQNIIIKDEK--TRAASIVDLSAGFNLIQ----- 1275
Db 217 DITVHLNGIGSLTDL--VGSPTHIDGG---DQSTHYTRAASIKDVLNAGNINIKGVKA 271
Qy 1276 -----GNGEAVDFVSTYDVFNFADGNATTAKVYDDTSKTSKVYDVNVDVDTTIEVKDKL 1331
Db 272 GSTTGQSENVDVHTYDVFELASADTETTVTVDSKENGKRTVEKIGAKTSVIKEDGKL 331
Qy 1332 GV-KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDITOTAKAGSQAANSAG 1390
Db 332 FTGKANKETNKVDGAN--ATEDADEGKGLVTAQDKVIDAVNKTGWRIKTTDANGONGDFAT 389
Qy 1391 YVD-----ADGN---KVYDSTDNKYYQAKNDGTVDKTEVAKDKLVAQAQTPDGTGLAQ 1441
Db 390 VASGNTVTFASGNTTATVTNGTDG--ITVKYDAKVGDGLKLDGDKIAA-----DTALT 442
Qy 1442 MNVKSVINKEQVNDANKKOGINEDNAPVKGLEKASDNK---TKNAAVTVGDLNVAQAQTP 1498
Db 443 VN-----DGKNANNPKGVADVASTDEKKLVTAKGLVTA--LNSLSWT 484
Qy 1499 LTFAGDTGT-----TAKKLGTELTIKGGQDTNKLTDNNIGVAGTGDGFTVKLAKDLT 1551
Db 485 TAAEADGDTLDGNASEQEVKAGDKVTEKAGK-----NLKVKQBGANFTSLQDALT 535
Qy 1552 NLNSV-----NAGGTTIKDDKGVSEFVSSGOA--KANTPVLSANGLDLGKGVISNVKGT 1603
Db 536 GLTSLTIGTGNGAKTEINKDGLIITPANGAGANNANTISTVKDGISAGSGSVKNVYVSG 595
Qy 1604 KDTDAANVQOL-----NEVRNLLGLGNAGNDN-----ADGNQVNIAD----- 1640


```
Db 596 KFGDANFDP L TSSADNLT KQNDAYKGLNLEKGD KDTKPTPVADNTAATVGD LRLGLW 655
QY 1641 -IRKDPNSGSS-----NRTV I KAGTVLGGKGNNDTEKLTATGGIGQVGVKDGNGANGDL 1692
Db 656 VISADKTTGSGTEYHDQVRNANEV---FKSGNGINVSQKTVNGRREITPEL---AKGEV 709
QY 1693 --SNVWKTKQKSGKALLAT---YNAAGTNYLTNNPABAI DRINEQGRFRHFVNDGNQ 1747
Db 710 VKSNEFTVKETNGKTSKLVKVGDKYYSKEDIDLTTGP-----KLKDN- 753
QY 1748 EPVVOGRNGIDSSASGKHSVAIGFAKADGEAAVAIGRTQAGNQSTAIAGDNAQATGDQS 1807
Db 754 -----TVAAYQDK-----GKVVSVVDTEATITNK 780
QY 1808 IATGTVNVAGKHSAGTDPSTVKADNSYSVGNNO-----FTDAITQDVFVGVGNITV 1861
Db 781 ---GSGVYTGQVADAI-----AKSGPELGLADEADAKRAFDDTKALSAGTTEIVNA 830
QY 1862 TESNSVALGNSAISAGTHAGTQAK----- 1886
Db 831 HDKVRFRANGLNTRKVSAAATVESTDANGDKVTTTFTVKTDELPLTQIYNTDANGKKITKVVK 890
QY 1887 -----KSDGTAGTTT-----AGATGTVKGFAGQTAVG 1914
Db 891 DGOTKVELNADGTADMTKEVTLGNDVSDGKVVKONDGKWKYHAKADGTADTKGVEVSD 950
QY 1915 AVSVG-----ASGAERRIQNVAAGEVSATSTDAVNSQLY---KATQSIANAT 1959
Db 951 KVSTDEKHVYVLPNDQSKGKGVVDNVANGDISATSDAINGSQLYAVAKGVYTNLAGOV 1010
QY 1960 NELDHRIHQENKANAGISSAMAMASQPAYIPGRSMVTGGIATHNGQGAVALGSLKSD 2019
Db 1011 NNLEGVKNGYKGRADAGTASALASQLPQATMPCKSMVAITAGSSYQCGQNGLAICVSRISD 1070
QY 2020 NGQWFKINGSADTQGHVGAAGAGFHF 2047
Db 1071 NGKVIIRLSTGTSQGTGVAAGVGQW 1098

RESULT 7
ID Q9F2D8 PRELIMINARY; PRT; 1107 AA.
AC Q9F2D8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE SAPB PROTEIN.
GN SAPB.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CF18;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ277623; CAC14217.1; -.
SQ SEQUENCE 1107 AA; 113223 MW; F5C7CA651FED51AB CRC64;

Query Match 6.4%; Score 661; DB 2; Length 1107;
Best Local Similarity 23.3%; Pred. No. 5.6e-20;
Matches 298; Conservative 195; Mismatches 418; Indels 368; Gaps 58;

QY 924 DALVNA-----KQIAEN---INTLAKEIHTTKGTA---DTALQTFYKVKVDENN 966
Db 42 DALANAGNDTGSINQNTDTATNTTSNNLSNVTTLTDDALLMDAASGTFASR---NG 98
QY 967 NADDANAITYGQRNANNQVNTLTKGEN--GLNKTDKNGTVFGINTTSLGKAGKSTLN 1024
Db 969 VNYSVLNLGDSGGTTRIGNVSAAVNDTDAVNAQLKRSEEAANTYTDQKMGEMNSKIKG 1028
```

```
Db 99 SASKITNLAAGTAAAD---STDVANGSQLFDTNEKVDQN-----TADITNTNNSIN 146
QY 1025 DGGLSIKNPTCSBQIQVAGDCVFAKVNNGVVGAGIDGTRITRDEIGTGTNGSLDKS 1084
Db 147 QNTTDIATNTS-----INNLS-----NSVTTLTDDALLMDAASGTFAS 186
QY 1085 KPHLSKDGGINAGKKITNIQSGETAQNASHDAVGTGGKIYDLKTELENKISSTAKTAQNSLH 1144
Db 187 R-----NGSASKITNLAAGTAAADSTDAVNGSQLFDTNEKVDQN--TADITNTNNSIN 237
QY 1145 EFSVADQGGNFTVSNPYSYDYTSKTSDDVI-----TF-AGENGITTKV-NKGVVRVGI 1195
Db 238 QNTT--DIANTNTSINNLSNVTTLTDDALLMDADSTFSASRNGSASKITNLAAGTAA 295
QY 1196 DOTKGLTTPKLTGVNNGNKGIVIDSQN--GQNT--ITGLSNTLANVTNDKGSVTRTEQONI 1252
Db 296 DSTDAV-----NGSQLYETNQKVDQNTSAIDINTSITNLSSD----- 333
QY 1253 IKDEDKTRAASIVDVL SAGNLOQNGEAVDFVSTYDVTNFDAGNATTAKVYDDTSTKTSK 1312
Db 334 -----NLSWN-----ETNFSFASHGSSSTTNKITNVAAGELSE 366
QY 1313 VYVDVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLT 1372
Db 367 ESTDA-VNGSQLFETNEKVDQNTTIDIAANTN-----ITQNSTALENNTS 411
QY 1373 SGTQITAKGSAQNNNSAGYVDADGNKVIYDSTDNKYIQAQNDGTVDKTEKVAQKLAQA 1432
Db 412 VSDINT-----SITGLTD--NALLWDE-DTGAFSANHGGSTKITNVA-----AGA 454
QY 1433 QTPDGTIAQMNKVSINKEQVNDANKKOGNEDNAFVKGLEKAASDNKTKNAAVTVGDNL 1492
Db 455 LSEDST-----DAVNGSQLYETNQK-----VDQNTS-----AIADIN 486
QY 1493 AVAQTPLTEAGDTTAKKLGETLTIKGGOTDTNKLTDNNITGVVAGTGDFTVKLAKDLTN 1552
Db 487 -----TSTINLTGTDALSWDDEGAFSASHGTSCTNKIIN-----VAAG-----ELASDT- 531
QY 1553 LNSVNAGGTIKIDDKGV---SFVDSGQAKANT--PVL SANGLDLGGKVISNVGKGTDTD 1607
Db 532 -DAIN--GSQLYETNMLISQYNESISOLAGDTSETYITENG--TGKVI RTNDNGLEGQD 586
QY 1608 A-----ANYQOLNEVRNLLGLGNAGNDNAGNQVNTADIKKDPNSGSSSRTVIKA--- 1658
Db 587 AYATNGATAVGYDAVASGAGCALGQNSSSSIEGSA-----LGSGSTSNRAITTGIRE 641
QY 1659 -----QTVLGGKGNNDTEKLTATGGIGQVGVKDGNGDLSNVVWKQKDGSKKALLATY 1712
Db 642 TSATSDGVI---GYNTDRELGLALSIGT--DGESTYQITNV-----ADGSE----- 684
QY 1713 NAAQTNYLTNNPABAI DRINEQGRFRHFVNDGNQEPVVOGRNGIDSSASGKHSVAIGFQ 1772
Db 685 ---AQDAVTVRQLQNAICAVTTTPTKYVHANSTEE-----DSLAVGTDLSLWAGAK 731
QY 1773 AKADGEAAVAIGRTQA---GNQSIAGDNRQAATGDSOIAIGTGNVNVAGKHSIGAIDPST 1829
Db 732 TIVNADAGIGIGLNTLVMDAINGIAIGSNARANHANSIAMNGSQTT---RGAQTDYTA 788
QY 1830 VKAD-----NSYSVGNNN---QFTD---ATQTDVFCVG-----NNITVTESN 1865
Db 789 YNMDTPONSVEFVSQVSGEDGQRTNVNVAAGSADTDVNVGQLKVTDAQVSRNTOSITNLN 848
QY 1866 -----SVALGNSAISAGT---HAGTQAKSKDQTA-GTTTTAGATGTVKGFAGQT 1911
Db 849 TVQSNLDRVTNIENGIGDIVTTGSKYFKTNTDGAADANAOGADVAIGSGSIAAENS 908
QY 1912 AVG-----AVSVCASGAERRIQNVAA----- 1932
Db 909 ALGNTSVADEANTVSVGSSQTOORRITNVAAGVNNNTDAVNAQLKASEAGSVRYETNADGS 968
QY 1933 -----GEVSA--TSTDVANGSOLYKATOSIANATN---ELDHRIHQ 1968
Db 969 VNYSVLNLGDSGGTTRIGNVSAAVNDTDAVNAQLKRSEEAANTYTDQKMGEMNSKIKG 1028
```


Thu Sep 13 14:19:01 2001

QY	1969	NENKANAGISSAMAMASPOAY	IPCRSMVTGGIATHRQGOAVAVGLSKLSDNQGWFKN	2028
Db	1029	TENKMSGGIASAMAMAGLP	POAYAFGANNTSIAGTGFNGESAIVAIGVSMVSESGGWYKLG	1088
QY	2029	GSADTQGHVGAAGAGFHP	2047	
Db	1089	GTSNSOGDYSAAGAGFOW	1107	
RESULT	8			
Q9F285		PRELIMINARY:	PRT:	3705 AA.
AC	Q9F285;			
DT	01-MAR-2001	(TremBLrel. 16, Created)		
DT	01-MAR-2001	(TremBLrel. 16, Last sequence update)		
DT	01-MAR-2001	(TremBLrel. 16, Last annotation update)		
DE	YAPH PROTEIN.			
GN	YAPH.			
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Yersinia.			
OX	NCBI_TaxID=632;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CO-92 BIOVAR ORIENTALIS;			
RA	Henderson I.R., Nataro J.P., Cappello R., Stein C.;			
RT	"Evolutionary origins of the autotransporter proteins.;"			
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AJ277631; CAC14227.1; -			
SQ	SEQUENCE 3705 AA; 370756 MW; 714PDF16455968C9 CRC64;			
Query Match	6.1%;	Score 627;	DB 2;	Length 3705;
Best Local Similarity	21.1%;	Pred. No. 7.4e-18;		
Matches 487;	Conservative 329;	Mismatches 876;	Indels 614;	Gaps 107;
QY	1	MNHYKVIENKATGTFMAVEYAKSH---	STGGGSCATG	-----36
Db	1	MNTFKVWNASLVNVVVSELAKRIKTKSSRNLLISEGLVPKFEQSMVSKLFRKNLLAL	60	
QY	37	QVGSVCTLSFARIAALAVLV-----	IGATLGSAYAQKDKTKH	-----AIGEQNPRR 85
Db	61	SLGSIVFLSTGPVPAADITVSTQAELSAALSNGTY-----	DKIILGADITLIGSLVNT	115
QY	86	SGTAKADGDAIRAIGENANAQGOAIGSSNKTV--NGSSLDKIGIDATGQESIAIGDV	144	
Db	116	SNQVVIDQGGFGLTVNNTNYGLVSSGSGTLTLQNSKID----	SANYYSMVVLNG--	169
QY	145	KASGDASTAI-----	GSDDLHLDDQH-----	NPKHPKGTLLINDLINGHAYLKEIRS 191
Db	170	-----ANTAVNVIYNNIDFLGSSQLIYMGAYGAATNSIMTFGDILND-----	VVWNDRA	218
QY	192	SKDNDVKYRRRTASCHASTAVGAMSYAOGHFSNAGFTRATAKSAYSLAVGLAATAEGOST	251	
Db	219	OBIGEVNKLATGRPHVTHTGSSVT-----	SPVSTGCGANTSTMDFAAGADVKIDRTGST--	273
QY	252	IAIGSDATSSSIALGALGAGTAAQLOGSIALGOGSVVTSQDNNSRPAYTPTNQALDKFQ	311	
Db	274	-----GDLTSGVNAFAY-----	TFADGASFELIANON-----	VFS 304
QY	312	ATNNTKAGPLSIGS--NSIKRKTIINVGACVKNYDAVNVAQLEAVVWKAKERRITFQG--DDN	369	
Db	305	GTTTNRG--LEIGSYNSID-----	GRGSGV-----	KIVLQSRSDG 337
QY	370	STDVKTIGLNTLTIKGAETNALDNNTI-----	GVKKEADNSGLVKYKLAATLNNLFEVN	423
Db	338	SIISGNGIDNATNAAGINNAGSDANVYNLGTGSILKATNTG--	ILATKNANNASDIY	395
QY	424	-----TTTLNATTIVKV-----	GSSSTTAELLSLTFQPTQNTSQSTSKTVY	467
Db	396	IRSAGDITAATGISATHNGTGVTKIKNDGTITSTTATGAISSASIKESIVDNTDGTITAT	455	

QY	468	GVNGVKFTNNAETTAAGT--	TRITRDKIGFARDGDVDEKQAPYLDKKQKLVGSVAITID	525
Db	456	AGTCVNVLASAILNFGGTINTSATANGITFA-----	GTEGGHTLTDLTINLLGTGIALS	510
QY	526	NGIDAGNKKISNLA-----	KGSSAND--AVTIEQLKAAPTLN---AGAGISVTPTEISV-	575
Db	511	N-VAGVNLTLSNVTLNLNGTALNSLTGLTLDVSLNGRNTINIEGAGIGIAATNTLWTF	569	
QY	576	DAKSGNVTAPTYNIGVKTTELNSDGTSDKFSVKSGSTNNSLVTAEHLASLYLNEVNRITADS	635	
Db	570	DAEALDINVNGAGIGIQAT-----	GGGVNLS-----ASNLIINVANTLGT	609
QY	636	ALQSFVTKEDDDDDANAITVAKDTTKNAGAVSILKLGKNGLTV-----	ATKK-----	683
Db	610	ALQI-----	TDGIDNTTIGNEIQLNAENATAINFGLSSSKTLLNNGTIGKSVIFAGVA	663
QY	684	-----	DGTVTFGLSDSGLTIGKSTLNNDLTVKDTNEIQIOVGANGIKFTNVNGS	733
Db	664	DHIINNGTLDGTLTTGAGNDT--	LVLDDSSOSNDVINLGDGNNSVTIQNGATVSSIIITGN	722
QY	734	NPGTGIANTARITRDKIGFAGSDGAVDTNKPYLDQDKLVQGVNVKIINTNGINAGGKAITGL	793	
Db	723	GNDTFTINGMSVGSTYLGSLDAGTGLNTXNXASTDELAATSLQGTNTINLVDSHITLV	782	
QY	794	SPTLPSIADOSSRNIELGNTIQDKKSNAASINDILANTGENLKNNNNPIDFVSTYD--	IVD 852	
Db	783	S-----	DDNIGSGMVNIDSSSELLFGSTFDGILH	811
QY	853	FANGNATTATVHTDPANKTSKVYDVNVDDTTHLTGTDDNKKLGVKTKLNTKSANGNT	912	
Db	812	ATLCAGTGSALVNNSAN-----	VSLQEASM--FAGTQWVNOGCGALTASNQLGSAKI	862
QY	913	ATNPNVNSDEEDALVN-----	AKDIAENLTLAKETHTTKGTADTALQTFVVKKYVDEN	965
Db	863	GIDGTLN--LDNTALFNHVLGTGLNLVAKNLATPAFDFGTVGGAFSGIVNLT--	KTTFA 919	
QY	966	NNADDANAI--TVGOKNANNQVNTL--	TLKGENGLMIKTDKNGTVTFGINTTSGLKACKST	1022
Db	920	LSADNAAALASATLKLSDPSVTTGTTDRTLHGLDL--	SGGLTFIDGAVPQSTSGVVT	976
QY	1023	LNDGGLSINKPTGSEQIOVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGNGS--	L 1081	
Db	977	VTD--	LALNSGT-----VN-----ITGSG-----	SWDNTDPLATNVSIL 1008
QY	1082	DKSKPHLSKDGINAGGKKTINTQSGEIAQNSHDVATGGKIYDLKTE--	LENKISSTAKTAQ	1140
Db	1009	EQDRAGSTLELINA-----	TNV--TGDI--DALDLVNGTAITSQTGVQSAIQGGSTVA	1060
QY	1141	NSLHFEFSADEQGNFTVSNPYSSYDTSKT-----	SDVITFAGENGITPK--VNKGVVR	1192
Db	1061	NAIHNYGLASSNSNG--	DSGLYVNYTLTSALELLADGADALLATESGLTANRVLNAELFG	1118
QY	1193	VGIDQTKGLTTPKLVGNNNGKIVIDSONGQNTITGLSNTLANVYNDKGVSRTEQONI	1252	
Db	1119	VG-----	GLVVDAQNG-----ALTLANGSNRYEGTITVTAGEL	1151
QY	1253	IKDEDKT--RAASIVDVLS--	AGFNLOGNGEAVDFVSTYDYNFADGNATTA-----	1300
Db	1152	ILGANGAFGQTSLLDITAGSASANINGYSQTGAVTNVGTITLGGSGVLTSGLLTNGGILD	1211	
QY	1301	-----	KVTYDDTSK-----TSKVYVDVNVDDTTIEVKDKKLIG-----	VKTTTLTSTG 1342
Db	1212	LTGGALNLTXXGASTVAGGLTGAGTLNNGNLVSAAANSGLSGQTHIADVASVLTDTG	1271	
QY	1343	T-----	GANKFALSNOATGDALVKASDIYVAHLNLTLSGDTQDTAKGASQANN	1387
Db	1272	TLGTSAVEVLGTNLNAGANA--	AMTNVLSGGDTINTNAAV-----TLSC-----	NN 1315
QY	1388	S-----	AGVVDADGNKVIYDSTDNKYQAKNDGTVDKTKVEAKDKLVAQAQPDGTLQAMNV	1444
Db	1316	SFSGAHOIGTDBELTV-----	QGSNLGASSATVNL-----GTLTSHLI	1354
QY	1445	KSVINKBQVNDANKKOGINEDNAFVKGLEKAASDNKT-----	KNAAVTVGDNLAV--	1494

Db 1355 LNVSESIANVLSGVAGTVD--IIGADTALTANNSGFLGOYALAGNSKLTVASTNNLG 1412
Qy 1495 AQTPLTTFAGDTGTTAKKLGELTIKG--GQTDNKLTDNNIGVVGAGTGGTFLAKDLTNL 1553
Db 1413 ASSSVALLAG-----AGDTLSLSGFNCITFGNSVTGS--GVLOQVTDAAEV---TLTSS 1458
Qy 1554 NSV--NAGGTIKDDKGVSEFVDSGQAKANTPVLSANGL-----DLGGKV----- 1595
Db 1459 NGVSNAYTIDIADATLNDLDDIA-----LFNHVLTGNGLLNVAKNDASTAFDFGVGGAFFS 1514
Qy 1596 -ISNVGKGTDTDAANVQOLNEV-----RNLGLG--NAGNDNADGNQVN 1637
Db 1515 GIVMLTWTTFALSADNAALARAFLKLSDSDSVTTVGATDRTLHGLDLNGGTLIFDGS--- 1571
Qy 1638 IADIKKDPNCGSSNNRTV---IKAGTV--LGGKNNDEKLTAT-----GGIQVG 1681
Db 1572 -----PPQSOAGVVTVDLALNSGTISITGACNWEHEPVPVPPNVSLLEQDRGDILLE 1625
Qy 1682 VDKDGNANGDLSNVVKTQDKGSKALLATYNAAGQNTYLTNNPAEADIRNEOGIRPFH 1741
Db 1626 LINAANYTGNANNLDLLV--DGT-----AITSQTGVESAIOQGGSTVANAIHNYGLTSSN 1679
Qy 1742 VNDGNQEPVVOGRNGIDSSAGKHSVAIG-----FOAKADGEAAVAIGRQTQAGN 1791
Db 1680 GNGSGLYVNTLSALELLANGANALLATESGLTANRVLNAELFGVGLVY--DAONGA 1737
Qy 1792 QSIAGDNAQATGQSAIGTGNVAGK---HSGAIGDSTPVKADNSYSVGNNNQETDA 1847
Db 1738 LTLANGNNRYE-----GTTVTAGELILCANGACFGTSLNLIASCASANINGY---- 1785
Qy 1848 TQTVDFVGNNTVITENSVALGNSAISAG--THAG-----TOAKKSDGTAGTTTTAG-- 1899
Db 1786 RQT-----VG-----AVTNSGAVTLNGGVLTLNGGLTDLTGALNLAAGSGSTVAGGL 1838
Qy 1900 -ATCTVAGFAGQAVGAVSVGCASCAERRIONVAAGEVSATSDAVNGSOLYKATQSIANA 1958
Db 1839 TGAGTLNMGDLVASATNSGLSG--OThIADVASVTLTGTLGTSAVEVIGTL----- 1891
Qy 1959 TNELDHRHQENKANAGISSAMAMAMPOQAYIPGRSNVTGGIATHNGQGVAVVGLSKLS 2018
Db 1892 -----NLGANAAMTNVLS-----GGGVINTNAAVTLCNNNSFS-GAHOIG 1931
Qy 2019 DNGQWFKINGSADTQGHVGAAGAG 2044
Db 1932 TDGELTV---GQASNLGASSATVNLG 1954

RESULT 9

Q92D91 PRELIMINARY; PRT: 2340 AA.
AC Q92D91;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CELL SURFACE ANTIGEN (SCA3).
GN RP451.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sierichitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
DR EMBL: AJ235271; CAAL4908.1;
SQ SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;

Query Match 5.6%; Score 582; DB 2; Length 2340;
Best Local Similarity 21.2%; Pred. No. 3e-16;
Matches 499; Conservative 313; Mismatches 879; Indels 658; Gaps 119;
Qy 25 SHSTGGSCATQGVSVCTLSFARIAALAVLIGATLGSAY-AOKKTKKHIAIGEONP 83
Db 252 AYDLGSSIIITGIDNIGSLD-----TINVLLGSATFNSTILKATNINLKH----- 297
Qy 84 RRSCTAKADGDRATAIGENANAGQQAIAIGSSNKTVNGSLDKIGTGDATGQESIAIGD 143
Db 298 -NTSTLND-DNIIVIG---NIKNNKDI--LNFKVHGTNLDNEMITPAQCT---HGT 347
Qy 144 VKASGDASIAIGSDDLHLL---DQHGKPKPKG-TLINDLINGHAVLK-----EIRSSKDN 195
Db 348 LNFEGNATLNGNINNLNITLKFSGHGKTLNLOGNTKVDNLVFAFDSVLDGSGTISVNGLLDT 407
Qy 196 D-VKYRRTTASGHASTAVGAMSYAQGHFSNAGFRATPAKASAYSLAVGLAATAEGOSTAI 254
Db 408 DCVTFNNSNVNG-GTLIINAKNTISAKULLNA--TKAKIQINANLTMN--HPSAGDISDIRI 463
Qy 255 GSDATSSSLGAIALGAGTQRAQLOGSI-ALGOGSVVVTQSDNNSRPAYTPTQALDPKFOAT 313
Db 464 -ADMTIYTIDA-----KGNVNLLNNNAKIIFEGADSMALINTGVTDRTFTIY 512
Qy 314 NN-----TKAGPLSIGSNSIKR--KIINYGAG---VNKTDA 344
Db 513 NNLQSGNDEYGVKIEAIKKVITIANOSGPTIGQDNTHRLKELIVEGAGDIIIDDTIF 572
Qy 345 VVNAQLEAVVWAKERRITFOGDDNSTDVKIGLDMNLTIKGAETNALTDNNIGVVRKAD 404
Db 573 TKLLSINSTGOITFRTLDLAGGN---IAFGKHGTLVNGVTGTSITSENNOGIL--TI 627
Qy 405 NSGLKVLKAKLNLNTEVNTTLNATTVKVSSSTTAELLSDSLSTTOPN----- 456
Db 628 NSG-----NITGVIGTNELGLKLVNIGADPVTCSANFASVALTNFSSVLIADG 677
Qy 457 ---TGSQSTKTVGVNGVKFTNNAETAAIGTTRITRDKIGFARDGVDDEKQAPYLDKK 513
Db 678 VTLGTEVTHNTKGV--LSLGTGSNITGQGTNSAALEKI----- 716
Qy 514 QLKVGSAVITDNGIDAGNKKISNLAKGSSANDAVTISQLKAAKPTLNAGAGISVTPTEI 573
Db 717 --NIGAGASNIDSIYAGSTVLTDTQTSLTNNDVVVNSNIITTAGNNSGK----- 765
Qy 574 SVDAKGNVTAPTNYIGVKTTELNSDGTSDKFSVKGSGTNNSLVTAHSLASVLYNVRFTA 633
Db 766 LIFTGNGGITG---NIGANGAALQ-----EVVFNGTN-----IGGTA 800
Qy 634 DSALQSTVTKBEDDDANAITVA-----KDT---TKNAGAVSILKLKGK----- 674
Db 801 NS--QNETVAH--SAANVTIGLTTGALKYKDTGTIAHGLGVGDIDFNKNKAGFILGD 855
Qy 675 -----NGLJVATKK---DGTVTFGLSODSGLTGTCKSTLNNNDGLTVKDTNEQIQVG 721
Db 856 GAMIDGSVLCNGGVAGTLDLFDGDNVTQNIQADNANSI--STINTQG---DNTKNVTI- 908
Qy 722 ANGKFTNVNGSNPCTGTANTARTDKIGFAGSDGAVDTNKPDLQDKQLQGVNKIYNT 781
Db 909 ANDIFVDNIHFTNGI--LQLGNIITPHNIDFGANGGTLEFN-----GN-----NT 952
Qy 782 GINAGGKAITGLSPTLPISADQSSRNIELGNTIQDKKSNAAASINDIILMTG-----FN 834
Db 953 -YLNALIVNGONGLLNAFTNLKASDDTIGTV-----KIINIGIGTGPQNET 998
Qy 835 LKNNNPIDFVSTYD-IVDFANGNATTATVTHDHTANKTSKVYVDVN-----VDDTTI 885
Db 999 IQVNNKNTLVSSVNSSINF--GDANSQLILSAPVDQTIKFINNLNETGGGIITLDSNGN 1056
Qy 886 HLTGTDGDNK-KIGVKTTKLNKTSANGN-TATN-----FNVNSS---DEDALVNAKD 931
Db 1057 NLTISGNGIILKSGKGNELSLNITKGVTVTVDLNDLQIHNQNTNNGALFDDQSLTSAK- 1115

1799 NAQATGDSATAGTGNVAG---KHSAGIDPSTVKAODNSYSGV-----NNQFTDA-- 1847
2099 NI-----DNSIVGAAYTMADSKVKHKNDKNGDRTRAKSNIYSIYGLYNWLTNNFEVEAIG 2154
1848 -----TOTDVFVGNNITVTESNVALGSNSAISAGTHAGT----- 1883
2155 VYGRNKKIKNYEKRTITTDQIAICKEINTFYSEYELGGYNYLIS---HRTTITPFGMRY 2211
1884 QAKSDGTAGTTTAGTGTGVKGAG--OTAVGAVSVGASGAERRIQNVAAAGEVSATSD 1941
2212 ATFKNGYKENNTTFQNLSEIKKNYDKFETILGLNSV-----THYLSQD 2255
1942 AVNGSQL-----YKATOSIANATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSM 1996
2256 IIKPELHWFYINYOCKNKLNPIDARLD-----GIDEPLTITIRKPAKITYN-- 2301
1997 VTGGIATHN 2005
2302 LGGGISTKN 2310

RESULT 10
Q9ZHL0
ID Q9ZHL0 PRELIMINARY; PRT; 4919 AA.
AC Q9ZHL0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE LARGE SUPERNATANT PROTEIN 2.
GN LSPA2.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000.
RX MEDLINE=99030326; PubMed=9811662;
RA Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
protein";
RL J. Bacteriol. 180:6013-6022(1998).
DR EMBL: AF057696; AAC79761.1;
SQ SEQUENCE 4919 AA; 542602 MW; 5779201455CA69A0 CRC64;

Query Match 5.5%; Score 570.5; DB 2; Length 4919;
Best Local Similarity 19.2%; Pred. No. 2.5e-15;
Matches 484; Conservative 353; Mismatches 914; Indels 771; Gaps 109;

QY 2 NHTYKVFENKATGTETMAVAEYAKSHSTGGSCATQGVGS-----VCT----- 43
Db 3 NKRYKLIFFSKVKNCLVPVAENIKSASGNSGSSNSKIAEDQEEEDPSLACSLSSSIH 62
QY 44 -----LSTFATAALAVIVIGATLGSAAVAKKDKTKHIAIGBONOPRRS- 86
Db 63 LGLRHNSPLKVFVKGNLSVLLSLMPAMQVWADSSNAIVDHSIGAKQTAVDERD-PKNGK 121
QY 87 -----GTAKADGDRAI-----AIGENANAQ-GQATIAIGSNKTVNGSSLD 126
Db 122 EKVIVINIAKPD-EOGISDNHFSKFNIPNSAVFNNSIKESNSQLVGLLGENKILGSAQA 180
QY 127 KIGTDATQGESIAGGVDVKAAGSASTAGSDLLHLLDQHGKPKHKTGLNDL--INGHA 184
Db 181 TIFNQVTGDOESKISGGLEVFEKA-----DLFII-----NPNGLVTLNGVKTINTDR 227
QY 185 VLKEIRSKNDVK-----YRRTTASGHASTAVGMSYAQGHFSNAFECTRATKAYSALV 240
Db 228 FVASTSEVPEPHIKQLNVQRKVIIGKGVATNGLS-----HFD----- 266
QY 241 GLAATAEQSTIATGSDATSSSLGAIAGTAGTAQLOGSIALGQGSVVTOSDNNRPAYT 300
Db 267 VVAKNIEQOGKVSIEGDSKPAKLANVTFAGNLTY-----DVTNRDV-N 309

932 IAEENLTLAKEITHTKTADTALQTFYVKKVDENNADANAIVTQ-----KNAN 982
1116 -IKNIN-----IGTVAGGA-----TITDAINDFDLNTSGWFKHQSILELKNSSNTN 1164
983 NQVNTLTLKAGENLN-----IKTDKNQTVTFGINTTSGL--KAKKSTLNDGG 1027
1165 DHTITLTSALDPGNNQFGIILITOTNKLITDNGNVAVTLGTANHMLKQLTFASIDNGA 1224
1028 LSTKNPTGSEQOVGADGVKFAKVNNGVVGAGIDGT--TRITRDEIGFTGTNGSLDKSKP 1086
1225 IALKVGINVENTLNTKDJLELVANVLFNKNTTYTATNGINGHVDFQGNAGVI----- 1279
1087 HLSKDGINAGGKKITIQSGEIAQASHDAVGTGKIYDLKTELENKI-----SSTAKTAONS 1142
1280 NLNDDEIDGVSYSTGVNGTGNFNGSGRVGT-----LINNVMLQAGAGDVLSLA 1330
1143 LHBFSVADEG-----NNPT--VSNPYSSYOTSKT-----SDV 1173
1331 SGNYSITEIQGNGNNLTFANSHLTTDINKTGGQDLNLVFNNGSVSGSIGANAAVGDI 1390
1174 ITFAG-----ENGITTKVKNKVVVRVIGID--QTKGLTTPKL-----T 1207
1391 IINAGSVNFNTLKSQNVISDQATMOVNNVTATDISKNANNGTLKLNHTPINITST 1450
1208 VGNNGKGIVIDSONGQNTIT-----GLSNTLANVTNDKGSVKT 1246
1451 LGNNAIG--TIEVANNDVTITGTLOAQNIHFSNATQAATLTLAGASQVNTNITAGNNIHT 1509
1247 TE-----QGN--IHKDEKTRAAISVDVLSAGFNLOGNG-----EAVDFVSTYDVTNPAFG 1295
1510 LEVTFDTGNDGIIGDAN--NRLKSI-----ELTNGTIVTINSHPVYSSITANNAAQ 1560
1296 NA-----TTAKVTYDTSKTSKVYDVNVVDDTTI--EYKDKKLGVKTTTLTSTGTGANKFAL 1350
1561 NVKLNIEGGITVDLGSIKSLANVOISEDTTIRGDVYSKYLVNIDACKTINFDRGDNMNP 1620
1351 SNOATGDALV-----KASDIVAHLNLTSGD--IOTAKGASOAN--NSAGYVDAD--GNKVIV 1401
1621 KNLDPDALIDLVLPSLSLNFNTFTDIKADNLNFADDTATANEKDAVVIDAHIDNGGIL 1680
1402 DSTONKY--YQAKNDGTVDKTEKAVOKLVAQAQTPDGTTLAQMVKV-----VI 1448
1681 KFNDAWLTQEIKNANII-----EIASDKF-----LLQKNIKAATLIADNANLVLL 1727
1449 NKEQVNDANKQGINEDNAFVKGLEKAAASDNKTKNAAVTVGDLNAVAQTPLTFAGDTGTT 1508
1728 DNVEVNTNLNVNRDIVLDLA---NYELKVTGNVTHNGLLTI-----ITY---FDTA 1771
1509 AKKLGETLTIKGQDRTNKLTDNNIGVVVAGTGTFTVKLAKDLTLNLS-----VNA 1558
1772 LQKGGHILVSQGSNVMSDL-DNLI-----IKIKASHDITNITSDBTKHQIYKLETGA 1822
1559 GGTKTDDKGVFVDSGQAKANTPVLNGL-----DLGKVISNVG-----KGTKDTDA 1609
1823 IYTPVPQTKVIIIDASEQNKVFKVWADANGLVLLTDTGCR--DGTGGRDDTGRGNTD-- 1878
1610 NVQOLNEVNLLGLGN-----AGNNDAGNOVNIADIKK--DPN-----S 1647
1879 -----NGCRDNCVGNISNNSSNEAGGSSSDKN-YGITDVVPFIDPSPILDYTKNNYVAS 1932
1648 GSSS-----NRTVTKAGTVLGGKGNNDTEKATGQIOGVVDKDG-----NANG-----DLSNV 1695
1933 GIANQLINHVKFGNTDAGKLLNDLGFMSPNRVTTETDRLSNRINVLNGLNEGVVGLNGI 1992
1696 WVK-----TQKD-GSKKALLA---TYNAAGQNTYLTNNPAPADIRINEQGITR 1738
1993 EVENFLTDIATNMDNFTAKEIGNRLELSDANTVNLKNTNTLLNKK-----IN 2041
1739 FFHVNDGNEBPVVOGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGROTQAGNQSIAGD 1798
2042 LKRLNTNNOAIIAAGDE--DNVITGWGMSFYGKIKONSKNS--ASGYQSNWTCGGIIGDPY 2098

Qy 301 PNTQALDPKQATNNTKAGPLSGSNS-----IKRKINVGAGVN-----KTDVAVN 346
Db 310 RNT---NPKRPITDNTKONIAISGESAGSMYGRNIFVTDKAGVNHQGVIFAEDDIN 366
Qy 347 V-----AQLEAVYKAKERRITFGQDDNSTDKVIGLDNTLTITKGGAETNALT 393
Db 367 ILTDGDSRLNKVYADVVRV--GKOTELANNGOIHA--DOOLILNATGHVKLNDGSSVIS 423
Qy 394 DNNIGV-----YKEADNSGLKYKLAKTLLNLTENVNTTLLNATVTVKVGSSSTTA 443
Db 424 NNLGISALNLTLENATVSANLSFRVNTDKLNLNLSKVSARAAD-----LQSGNLNLDKA 479
Qy 444 ELLSDSLTFQPTNGSOSTSKTVYGVNGVXF-----TNAETTAAGITRITRDKIGF 496
Db 480 SVLAHKLTLLNISDVSLN--NOSKLSANLTKVKVRLDNLNANSELSA-----NNLT 529
Qy 497 ARDGDVDEKQAPYLDKOLKQVGAITDNGID-----AGKKISNLAKGSSANDAVT 549
Db 530 NTSNNITLK-----NKSKEFTAGNMTLVNTNVTLNNDSELANLTLNVTKNVTLND-- 582
Qy 550 IEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAPTYNIG-----VKTTELSNDSGTS 603
Db 583 -SKLSANKLDLN-----VTDNVTLSKS-----TSLAGELTFKKVKVNTLNDSE-- 626
Qy 604 KFSVKGSGTNSLVTAEHLASYLEVN---RTADSALQSFTVKEEDDDANAITVAKDTT 660
Db 627 -----LAANNLSLNASHNVTLNKSKLSAQKADIKAVNLTLNDTTELAKNLDINSTTI 680
Qy 661 KNAGAVSIL-----KLKCK-----NGLTVATKKDGTVTGLS 692
Db 681 TNGTIAGIFANITTEKLNKKEKALIAEONLNTVNGSHYENKGDIVSKDATTFSKN 740
Qy 693 QDSGLTICKSLNLDGLTVKDTNQIANGIKFTVNGSNPGTGIANTRIDRIGF 752
Db 741 SD-----FTSNGSKLVNAQOLKAVNNFTISO-----GDDTLIGNVTLSNGT 785
Qy 753 AGSGAVDTNKPYLDDQKLVGNV--ITNTGINAGGKAITGLSPTLPSIADQSSRNIELG 811
Db 786 FTNSGNLTUVK-----TLDVGDIQNFNKG-----NLTVG 815
Qy 812 NTIQDKKS-----NAAINDILNTGF-----NLKNNNPIDF 844
Db 816 EDLHKSKTKITNDGKLISLKNLISSEADFINGTLLGTEALKIATKGNFTNEKAILA 875
Qy 845 VSTYDIDVFANGNATTATVTHDTA---NKTSKVYDVVNDVDTTH-----LTGTDN 893
Db 876 SNLLDISVAEGKKTENNFTIESGKNLNTNTGAF--LNVDNATIRSEGLNITSTGVSN 934
Qy 894 KKLGVTKTKLNKTSANGNTATNFVNSDDEALVNAKDIAENLNTLAKEIHTTKGTADTA 953
Db 935 NGTLISNERLNTS-----AANF--TNEISNGTVMISNGL-----LNIAKQGNITNKNLIAS 983
Qy 954 LOTFTVKKVDEN--NNADDANAITVGOK-----NANNVNTLTKE-----NLNLKTD 1001
Db 984 RQQLNLTAVDNITNDSNISKIAVLISLGNLSNLSKQVYNL---GEIYAGNINISVAH 1040
Qy 1002 --KNGTFTGINTTSGLKAGKSTLNDGGLSINKNPTGSEQIOVGADGVKFAKVNNGVYGA 1059
Db 1041 QLKNDVKLMGDIITK-----TKGOASYKLYQASNGHFGNDGSSGYSEGDLNKGK 1092
Qy 1060 GIDGTTTRITREIG--FTGTNGSLDKSKPHLSKDGINAGK-----KIT--- 1101
Db 1093 FADLDNKLTVQIRIGKIVAGRODLTENS-----NAGGKSEIINRGITNVNKNLSYDS 1143
Qy 1102 -----NIQSGEIAQNSHDAVTGKGIYDLKTELENKISSAKTQNSLHESVADQGN 1155
Db 1144 DVSENNMQSOKV-----DLYT--KIFEAKSDIE-----LTFKT--NGTHPVYL-----N 1184
Qy 1156 FTVSNPYSSTYDTSKTSOVIITPAGE---NGIITKVNKGVVVRGIDOTKGLTTP---KLTVG 1209
Db 1185 FKSNNKERYNSENTRNFKSIGDLNEALSAPAEIAEAYSSSSNYINPVSYLAALG 1244

Qy 2023 WV 2024

Qy 1210 NNNGKGIVIDSQONQNTITGLSNTLAN-----VTNDKSGSVRTE 1248
Db 1245 NAN-----NSSPHYLNTALKHILGNQDDLLKQENIKVLKOKWEDFKDKGASKMLD 1298
Qy 1249 QGNIIKDEDKTPRAASIVDVLSSAGFNLOGNEBPAVDFSVTYDVNFAAG--NATTAKVTYD 1305
Db 1299 ---LYPNTDEKAKIFAGIIRNGNDTISDVESEDFKKYS--KFQNGEAWAKNDGTDSYD 1353
Qy 1306 DTSKTSVVYDVNVDDTITIEYKDKLGVKVTTLTSTGTGANKFALSNOAGDALVKASDI 1365
Db 1354 STKASEYKVENVDHKE--NIDEHKLNGKHEITVPGVSFEN--LNNKNMHDQPKLGEI 1410
Qy 1366 VAHL--NTLSGDIOTAKGASQ-----ANNSAGVVDADGNKVIYDSTDN---KYIOAK 1412
Db 1411 DKSTIISLLAOPVYTEKSAARDSPRVQNQDKEALDNLRYTRLSYINQNNYLKAFYFNQ 1470
Qy 1413 NDTGVDTKTEKAKDKLVAQAOTPDGTLAQMNVKSVINKEQVNDANKKOCINEDNAFVKGL 1472
Db 1471 LDTEDDKLKGIKR-----IGDNYFEHOLITRIEKVADNHLTLKHLGH--DIALVKKL 1521
Qy 1473 EKAASDNKTKNAAVTVGDLNVAOQPIITFAGDTGTTAKKLGELT----- 1517
Db 1522 IDSAS-----IQAKDLNL-----KVGEALTKEQKDLNKEDIVWVY 1556
Qy 1518 -----IKGQTDNTKLTNNIGVWAG--TDGF 1542
Db 1557 KTEVNGEVLVPVYLAKQTIEVEKQGVGTQIRAGIIDVKVDDVRNTGTIAGYAVGL 1616
Qy 1543 TVK-----LAKDLTLNLSVNAAGTKIDDKGVSFVDSG----- 1575
Db 1617 EAKNKLKNTGILSQRUSKL-----VGKKGLESTGVTVYVDEGTATKVKRKARIKSEGHILE 1672
Qy 1576 -----QAKANTPVLSANGLDLG-----GKVISNVKGGTKFTDIDA 1608
Db 1673 TDKDKNVDLTASELKGNTGQIKAKDLNLDIYETSYKYKYEKLFCKNGGEIGDRVTOISQ 1732
Qy 1609 ANVOQLNEVRNLLGLNAGNDNADGNQV-----NTADIKKDPNSGSSNR-----TVIKAG 1659
Db 1733 AKSVGTDAFSDPHLHLSLEGDVNOTGSLNKANRTTGVVKGDFNTRKAGDLFHRQIDVTISG 1792
Qy 1660 TV-----LGGKNNDEKLTATGGIOGVDBK---DGNANGDLSNVWVKTK----- 1701
Db 1793 TVYSASASGGQSAGISLTDQGVETYNKTATAGANADVTNPMKRTETETSLTHRNSEF 1852
Qy 1702 -----DCSKKALL-----ATYNAAGQ 1717
Db 1853 NALSGELVVMGKADIGGVVDINRDVEV IKTPBEIAAEQKAAEEAKKAEVKEASEATAAKE 1912
Qy 1718 TNYLTNPAEALDR-----INEQGIREFHVNDGNQBPVVOGRNGIDSSA 1761
Db 1913 TEEAENDNVABEKDKTKPKFKLTDBEIAAAFETKGEDEFAAYKAREEE--DRKKGFTLSA 1970
Qy 1762 SGKHS-----VAIGFOAKADGEAAVAIGRO-----TOAG---NOSIAG 1797
Db 1971 BQIESTKARDEKETYYIELKVGVGAEAAHSAADAISNKARQIIDTQNGLKQDQTVLQ 2030
Qy 1798 DNAQ-----ATGDQSTAICTGNVWAGKHSAGICDPSVTK-----ADNYSYVGNNOFTDAT 1848
Db 2031 EASDVNLATGD-----LAGASAKLKFELSTIEKKSRSAGSDGRSLGRLNLAAR 2080
Qy 1849 QTDVFGVGNNTVTESNSVALGSNSAISAGTHAGTQAKKSDGTACTTTTAGATGVTKGFA 1908
Db 2081 GGDI--TLNNYETTESHLKARDNVNNSGV--TEQKDESNQSLKVTAGASSCGGMA 2137
Qy 1909 GOTAVGAVSVCASCAERRIQNVAAGEVSATSTDAVNGSQL--YKATQ--SIANATNELDHR 1965
Db 2138 GGCSAG--VSAGVSGS-----YNESNTFESTSHTNLSLRGSLRVEAGKDFNLSSNVVDVH- 2191
Qy 1966 IHQENKANAGISSAMAMASMPQAYIPGRSMVTVGGIATHNG---OCAVAVGLSKLSDNGQ 2022
Db 2192 LHL--DVKGDTNVVSKQDSYSRKERGVSYSVAGVGVSTAGGARPNGSVGLGVSAENENSK 2250
Qy 2023 WV 2024

Db	2251	IV	2252	
RESULT	11			
Q52657		PRELIMINARY:	PRT:	2021 AA.
ID	Q52657			
AC	Q52657			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)		
DE	190-KDA ANTIGEN (ROMPA).			
OS	Rickettsia conorii.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiae; Rickettsia.			
OX	NCBI_TaxID=781;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MALISH 7;			
RX	MEDLINE=94171067; PubMed=8125327;			
RA	Croquet-Valdes P.A., Weiss K., Walker D.H.;			
RT	"Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia			
RT	conorii (Malish 7 strain).";			
RL	Gene 140:115-119(1994).			
DR	EMBL; U01028; AAL17405.1.			
SQ	SEQUENCE 2021 AA: 203366 MW; BE943FBA3BDD5C43 CRC64;			
Query Match 5.5%; Score 567.5; DB 2; Length 2021;				
Best Local Similarity 23.6%; Pred. No. 1e-15;				
Matches 472; Conservative 198; Mismatches 745; Indels 589; Gaps 104;				
Qy	101	ENANAGQQAIAIGSSNKTVNGSSLDKIGTDGQESIAIGDVKASGDASIAIGSDDLH	160	
Db	10	KKAIQOGLKAALFTTAAIMLSSGALG-----IAGVGIAIYNNNAAPS-----	54	
Qy	161	LDQHGPKH-----PKGLINDLINGHAYLKEIRSSKNDNVKRYRTTASGHASTAVG	213	
Db	55	--DVGNNWNETAAGVANGTPARGPQNNNAFTYGGDYTVTADVADHIIATINAVDTPPI	112	
Qy	214	AMSYAQGHFSNAPGTRATAKAYSIAV--GLAATAGQSTIAI--GSDATP-----SSLAGAI	266	
Db	113	GLNIAQ---NTVVGSIITGGNLLPVTINAGKSLTLNGNDAANHGFDADADNYTGLGNI	169	
Qy	267	ALGAGTQAQLOGSIALGOGSVVVTQSDN--NSRPATPTPNTQALDPKFOATNNKAGPLSIGS	325	
Db	170	ALGGANAALIIQSAAPAK---ITLAGNINGGGIITVKTD-----AINGT-----IG-	213	
Qy	326	NSIKRKIIIVGAGVKNKDANVVAQLEAVVQWAKERRITFGQDNDSTVDKIGLDNTLTIGK	385	
Db	214	NTNATATNVGAG-----IATLEGAIKATTTKLT-----NAASVL-----TLTNV	255	
Qy	386	GAFNALDNNIGVVKADNSGLKVKAKTLNNLTENVNTTLNATTTVKVGSST-----	441	
Db	256	AVLTGAI-DNTTGV-----DNVGV-LNLGALSQVTVG-NIGNTNALATISVGAGKATLGA	308	
Qy	442	---TAEILSD---SITFQPNWGSGSTKTVYGVNG-VKFTNNAETAAIGTTRITRDK	493	
Db	309	VIKATTTKLTDNASAVTFTNPVVVTGAIDNTGNANNGIVTFTGDSVTGNTGNAL-----	365	
Qy	494	IGFARGDGVDEKAPYLDKKQKLVGSVAITIDNGIDAGNKKISNLAKGSSA-----NDAVT	549	
Db	366	-----ATNVWAGLILQOVGGVVKAN--TINLTDNASAVTFTNPVV	404	
Qy	550	IEQLKAAPLTLNAGAGISVTPTEISVDKSGNVTA-PTVNIQVKTTELNSDGTSDKFSVK	608	
Db	405	---TGAIDNTGNANNGIVTFTGDSVTGNTGNALATISVGAGKATLG--GAIKATTT	459	
Qy	609	SGSTNNSLVTAHSLASVILNEVNRDTASALQSFVKEEDDDANAITVAKDTTKNAGAVSI	668	
Db	460	KLTDNASAVT-----FTNPV-----VVTGAIDNTGNA-----	486	
Qy	669	LKLGKNGLIVATKKDGTGTGFLSQDSGLT-----IGKSTLNDGLTVKDTNEQIQVGAN	723	

Db	487	-----NNGIVTFT-GDSTVTGNTGNALATISVGAGKATLG--GAIKATTTKLTDNAS	538	
Qy	724	GIKFTN---VNGSNPCTGIANTRITROKIGFAGSDGAVDNKPYLDQDKLQVGNVKITN	780	
Db	539	AVTFTNPVVVTGAIDNTGNANNGIVT-----FTG-DSTVTGN-----IGNTALA	582	
Qy	781	TGINAGGKALTG---LSPITLPSIADSSSRNIELGNTIQDKDSNAASI---NDILNTGF-	833	
Db	583	TISVGAGKATLGGAIKATTTKLT-----NASAVTFTNPVVVTGAI	624	
Qy	834	NLKNNNPIDEVSTYDIDVFDFANGNATATTATVHTDANKTSKVYVDVNVDDTTIHLTGDD	892	
Db	625	DNTGNAN-----GIVTFTGDSVTGNTGNALAT-----VNVGAGTATLEGAI-	669	
Qy	893	NKILGVKTTKLNTKSANGNTATNFVNSDEDAVNAKDIAE-----NLNTLAKEIHFTVG	948	
Db	670	-----IKATTTKLTNA-ASVLTITNVNAVLTGAIDNTTGVNDVGNLNGALSQVTVGNI	723	
Qy	949	TADTALQTFVK-----KVDENNADDANAIV-----GOKNANNQVN	986	
Db	724	NTN-ALATISVGAGKATLGGAVIKATTTKLTDNASAVTFTNPVVVTGAIDNTGNANNGIA	782	
Qy	987	TLTLKGENGLNIKDKNGTVTFGINTTSGLK-----AGKSTLNDGGLSIKNPTGSSOIQV	1041	
Db	783	TFT-----GDSVTGIDIGNTALATISVGAGKATL--GGAIKATT--TKLTD	826	
Qy	1042	GADGVYKFAKVNNGVVGAGIDGTTTRITRDEIGFTGCTNGSLDKSKPHLSKDGINAGGKIT	1101	
Db	827	NASAVTF---TNPVVVTGAIDNTGNANNGIVTFTG-----DSTVTGNTGNT	869	
Qy	1102	NIQSEIAQNSHDVATGKGIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNFTVSNP	1161	
Db	870	NALA-TISVGAGKATLGGAIK-----ATTTKLTYN-----	899	
Qy	1162	YSSYDTSKTSVDITFAGENGITTKVNGVVRVGVGDQKGLTTPKLTGVN--NNGKIVIDS	1220	
Db	900	-----ASAVTF-----TNNVVVTGAIDNT-----GNANNG-----IVT	927	
Qy	1221	QNGQNTITG-LSNTLIANTVNDKGSVRTTEQGNIKDEDKTRAASIVD-----VLSAGFNL	1274	
Db	928	FTGDSVTGNTGNALATNVVGVAGTVLQAGSLD-----ANNIDFGARSTLEFNGPL	980	
Qy	1275	QNGEAVDFEVSTYDVFNFADGNATTAKYTYDDTSKTSKVYVDVNVDDTTIEVKKKLGK	1334	
Db	981	DGGKRAIPY---YFKGAIANGN-----NAILNVNPKSLTA-----SHLTIG	1018	
Qy	1335	TTTLSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDI---QTAGKASQANNSAG	1390	
Db	1019	TVAEINIGAG-NLFAIDASAGDVTILNAQDI--HFRALDSALVLSNLGTGVVNNILLAA	1075	
Qy	1391	YV--DADGKNKVIYDS-----TDNKYQAKNDGVYDKTEVAKDKLVAAQOTPDGTLAQ	1441	
Db	1076	LVAPGVDEGTVFDDGGVNGLNGISNVAGAARNIGDVGNK-----	1115	
Qy	1442	MYKSVINKEQVNDANKKOG-----INEDNAFKGLEKAASDNKTKNAAVTV---GDIN	1492	
Db	1116	FNTLLIYNAVYITDDNLEGTQNVNLNNADFTSSTAFNACTIQTNDATYTTIDANNGLN	1175	
Qy	1493	-----VAQOTPLTAFAGDTGTTAKKLGEBTLIKGGQTDTKNLTNNNGVAGTGDFTV	1544	
Db	1176	IPAGNIQFAHADAQLILQNSSGN-----DRTITL-GANIDPD---NDDEGIVILN---SV	1223	
Qy	1545	KLAKDLNLSNVNAGGT-KIDD---KGVSPVDSGQAKANTPVLSANGLDLGGKVTSNV	1600	
Db	1224	TAGKKLTAGGKTFFGAHKLODIVFKGEGDFGTAGTFTNTIV---LDITGQL--ELG	1277	
Qy	1601	KGTKDT---DAANVOQLNEVRNLLGLGNAGNDNAD--GNOVNIADIKKDPNCGSSSNRTV	1655	
Db	1278	ATTANVFLKDAVOLQTQGNIGGLDFD--NAKNGVTVLNNNVNVAGTVK--NTGCTNNGTL	1334	
Qy	1656	IKAGTVLGGKGNNDTEKLA---TGGIOGVYDKDGNAN-GDLSNVWVVKTDGSKKALLAT	1711	

Db	1080	-NTLGLVTABNAXIGTVAQINI--QDNKTFAINVKNADIEIL-----NAQAIDPKGA	1128
Qy	1287	YDTVNFADGNATKATVYDDTSKTSKVYDVNVDDTTIEVKDKKLGKVTTLTSTGTGCAN	1346
Db	1129	NSKLFVLNNSAT-----DDRVVTIKN-----DLPAFATGGG	1159
Qy	1347	KFALSNOATGDALVKASDVAIHNTLSGDIQTAKGASQONNSAGVYDADGN---KVIYDS	1403
Db	1160	MLLFG-----TANLITLQGD-----GAXKAIGTAGKXLASNLVLGKVAFNN	1201
Qy	1404	TDNKYIQAKNDGIVDKTEKAVAKDKLVAQAOTPDGTLAQMNVKSVINKQVNDANKKOGIN	1463
Db	1202	IDTNNVLAFF-----IINTQFVD-----	1220
Qy	1464	EDNAFVKGLFKAASDNKNTNAAVTVGDLNAVAQTPLTFAGDTGTAKKLGELFTLTGQGT	1523
Db	1221	-----VGGITNQIS-----VINIGAANA-----SLGLPAAAGSYTIDANG--	1256
Qy	1524	DTNKLTDNNIGVVGAGTDGFTVKLAKDLTNLNSVN--AGCTKI-----DDKGVSEF	1571
Db	1257	-----NVGILA--NGQTFINFAHEDAEVLVIONSAGNSTITLNAALDPGAPNKGKLA	1306
Qy	1572	DSSGOAKNTPVLSANGLDLGG-----KVTSNVKGT--KDTDA-ANVOOL-----	1614
Db	1307	D-SCXAGRVILASVRNATYGTAVNKLKLEFRGNGTFOQIDTDFANLELPVPAITYN	1365
Qy	1615	NEVRNLGLGNA-----GNDNAD--GNQVNTADIKKDPNSGSSNRTVIRKAGTVLGGK	1665
Db	1366	KDINKNLSFGAATALTQNGINGVNDPNNQAAVITL-----GANKNIT-----	1415
Qy	1666	GNNDTEKATGIGVGVYDKDGNANGDLSNVVVK-----NGHTINAMLVKGAGAVSITQGGNTSITEIQNGTAL	1466
Db	1416	GVNGT-TIATGASTI-----NGHTINAMLVKGAGAVSITQGGNTSITEIQNGTAL	1466
Qy	1709	L---ATYNAAGOTNVLNNPAEADIRNEQGIREFHVNDQNPVQVGRNGCIDSSASGKH	1765
Db	1467	LTLPANFNLTG-----SINKTGGQALKNFETNGSGVRGV-----	1501
Qy	1766	SVAGFOAKADGEAAVATGROTOAGNOSIAIGDNAQAT-----GDQSTA---ICTGNVAVG	1818
Db	1502	-----GTANSVGDITTYTGTTFNFASSVNAKGAATLCGTTSPADFTNTGAVTLA	1550
Qy	1819	KHSGAIGDPSTVKADNSYSGVNNQFDTATQDVFVGNNITVTESVAL-----GSNS	1873
Db	1551	KAS-----ITNFAKNVTATSF--AANNATINFGNSLAFNSNITGSGT	1590
Qy	1874	ALSAG-----THAGT-----QAKKSDG-----TAGTTTTAGATGT-----	1903
Db	1591	TLTLGASQVITYGTGSGFTDTLTNTTFDCAAKSDGHILIKSGSTLDLGSVSTLALVVTAT	1650
Qy	1904	---VKGFAGQTAAGVAVSGASGAERRIQNVAAGEVSATSTD-----AVNGSOLYKATQ	1953
Db	1651	NFDINNISPNTKYTVISAETAGG---LKPTPAGNVKVTYVNDNRNVNFTFDESTLTLPFAE	1707
Qy	1954	STANATNLDHR-----IHQENKANAGISSAMAMSPQAVIPGRSMV-----	1997
Db	1708	DIABEEVIEDFPKGPPLANIPNAAV-----KKSLEMEQAPIGSDACIAFDNWKGLNPIQ	1763
Qy	1998	TGCIATHNQGVAVAGLSKLSLSDNQWVFK-----INGSADTQGHVGAAGVAG	2044
Db	1764	ARDAVTHLQDVVKPSPDTIAANNQVILSNISSSLINLARNDAIQAPAAVAG	1816
RESULT	13		
Q9VTK8		PRELIMINARY; PRT; 2586 AA.	
AC	Q9VTK8		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DE	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)	
GN	CG18331	PROTEIN.	
OS	Drosophila melanogaster (Fruit fly).		

Db 279 TSSSDPVVEVAQGS--SSNGDGNSTQSSSTTTTTTTTTSSDGGQSTTLSDPVVEVSQGTNG 337
Qy 319 GPLSTGNSIKRKILINVAGVYKNTDAVNAQLEAVVAKWAKERRITFQGDNDSTOVKIGLD 378
Db 338 GNSSTQSSSTTTTTSSDEGQTTSSDPVGEV-----AQGSSNGDGNSTQSSSTTTT 389
Qy 379 NTLIKGAETNALTDNNIGVVKENDSGLKVLAKTLNLTNTVNTTLNATT-----432
Db 390 TTTTTSSDGGQSTTSSDPVVEVSQGTNG-----NSSTQSSATTTTTSSDEGQTT 440
Qy 433 -----VKVGSST-----STTAELLSLSLFTOPNTGQSQSTS-----KTVYGVNGVKF 474
Db 441 SSSDPVVEVAQGSSSIGDGNSTQSSSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNG---497
Qy 475 TNNATTAAGTTRITRDKIGFARDGVDDEKQAPYLDKKQLKVGSAITIDNGIDAGNKK 534
Db 498 -GNSSTQSSSTTTTTSS-----DEGQTT-----TSSSDP 525
Qy 535 ISNLAKGSSAN--DAVTIEOLKAAPTLNAGAGISVTPTEISVDAKSGNVTAPTYNLGV 591
Db 526 VVEVAQGSSSNGDGNSTQSSSTTTTTTTTTSSDGGQSTTSSDPVVEASQGT-----NGGN 579
Qy 592 KTTIELNSDCTSDKFSVKSGSTNNLSVTAEE--HLASYLNEVNRKTADSAEQSFTVKEEDDD 649
Db 580 SSTQSSSTTTTTSSDEGQTTSSDPVVEVAQGSSSIGDGNSTQSSSTTTTTTTTTSSDGG 639
Qy 650 ANAITVAKDTTKNACAVSILKKGKNGILTATKKDGTVTFGLSDGSLTGKSTLNNDGL 709
Db 640 QS--TTSSDPVVEAS-----QGTNGGNSSTQSSSTTTTTTTTTSSDGGQTTSSS 683
Qy 710 TVKDTNEIOVGANGIKFTNVNGSNPGTGIANTRARIDKIGFAGSDGAVDT--NKPYLD 767
Db 684 ---DPVVEVAQGSST-----NGDGNSTQSSSTTTTTTTTT--TSSDGGQSTTLSDPVVE 730
Qy 768 -----QDKLVQGVNKITWGINAGKALITGLSPLTSDOSSRIELGNTIQQDKK 819
Db 731 QSQGTNGGNSSTQSSSTTTTTSSDEG--PTTSSDPVVEVAQGSSSNGD--GNSQSTT 787
Qy 820 SNAASINDILNTGNLKNNNPIDFVS---TYDIVDFANGNATTATVTHDTANKTSK--V 874
Db 788 TTTTTTSS--SDGGSSTSSDPVVEVSQGTNGDGNSTQSSSTTTTTSSDGGQTTSSSDP 845
Qy 875 VYDV-----NVDDTTIHLTGDDNKKLGKVTIKLNTKSANGNTATNFVNSSD 922
Db 846 VSEVAQGSSTGDNSTQSSSTTTTTTTDIFRCPQRRTTTTTTTTTSSDGGESTTL----SD 901
Qy 923 EDALVNARDIAENLTLAKEIHLTKGTADTALQTFVKKKVDENNADANALITVGOKNAN 982
Db 902 PVVEVSQGTNGGNSSTQSSSTTTTTSSDEGQTTSSSDPVVEVAQGSSSN-----GDGNS 957
Qy 983 NOVNTLTL-----KGENGLNIKTDKNGTVTFGINTTSG-----1015
Db 958 QSSTTTTTTTTTTTTSSDGGESTTSSDPVVEVSQGTNGDGNSTQSSSTTTTTSSDGGQATSS 1017
Qy 1016 -----LKAGKSTLNDGGLSKNPTGSRQIQVGADGVKFAKVNNGV-VGAGIDG-----1063
Db 1018 SAPVVDISQGSSSNGDGNSTQSSSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNGGNSST 1077
Qy 1064 -----TTRITRDEIGFTGNGSLDKS-----KPHLSKD 1091
Db 1078 QSSSSTTTTTSSDEGQTTSSFGSCRSRSHRILFOCPRORQOLRRHLPMVANPPHLT--1135
Qy 1092 GINAGGKKTNIQSGEIAQNSHDAVTCGIYDLKTELENKISSTAKTAQNSLHEFSVADE 1151
Db 1136 -----PVVEVSQGTNGDGNSTQSSSTTTTTTTTTSSDGGQTTSSAPVSSSTTTTTTSS 1188
Qy 1152 QGNFTVS-----NPPSSYDTSKTSVDITFAGENGITTKVYKGVYRVGIDQTKG 1200
Db 1189 DGDQSTTSSDPVVEVSQGTNGDGNSTQSSSTTTTTTTTTSSDGGQTTSSAPVVEV-----TQC 1244
Qy 1201 LTTPKLTGVNNNGKIVIDSGONTITGLSNTLANVTNDKGSVRTTEQGNLIK-----1254
Db 1245 -----SSSNG-----DGNSTQSSSTTTTTTTTTSSDGGESTTSSDPVVEVSQGTN 1289

Qy 1255 -DEKTRAASIVDVLSAGFNLOQNGEAVDFVS--TYDTVNFADGNATTAKVYVDDTSKTSK 1312
Db 1289 GDNSTQSSSTTTTTSSDEGQTTSSAPVVEVTQSSSNGDGNSTQSSSTTTTTTTTTTSS 1348
Qy 1313 VVYD--VNVDDTTIEVKDKLGVKTTTLTSTGTGANKFALSNOATGD-----ALVKASD 1364
Db 1349 DGGESTTSSDPVVEVSQGTNGDGNSTQSSSTTTTTSSDGGQTTSSSGPVVDISOGSSSN 1408
Qy 1365 IVAHLNTLSGDT-----OTAKGASQANNS 1388
Db 1409 VLDNNDNNYDDIFRWANPPVLTPLWKSVEQMAITAPLSLRPPQOQTSSDEGQTTSS 1468
Qy 1389 AGYV-----DADGNKYVDSTDNKYYQAKNDGTVDKTKVEAKDKLVAQAQPDG--1437
Db 1469 SAPVVDISQGSSSNGDGNSTQSSSTTTTTTTTTSSDGGESTT---LSDPVVEVSQGTNGDN 1525
Qy 1438 --TLAQMNVKSVINKEOVNDANKKOGINEDNAFVGLKLEAKASDNKTKNAATV-----1488
Db 1526 NSTQSSSTTTTTSSDGGQTTSSSDPVVE---VAQGSSSNGDGNSTQSSSTTTTTTTTTTSS 1582
Qy 1489 -GDLNVAQOTPLTFA-----GDTGTTAKKLGETLTIKGGOTDNLKLTNNIGVWAGT---1539
Db 1583 DGGESTTSSDPVVEVSQGTNGDGNSTQSSSTTTTTSSDGGQTTSSSDPVVEVAQGSSSN 1642
Qy 1540 -DGFTVKLAKDLTLNLSNVNAG---TKIDDKGVSVFVDSGG-----1575
Db 1643 DGNSTQSSSTTTTTTTTTSSDGGESTTSSDPVELRDLPLVMVMTLPSRORQOLRRHL 1702
Qy 1576 -----QAKANTPVLS-----ANGLDL 1591
Db 1703 PMVANPPHLTPLMKSVKEQTAITAPLSLRPPQORHLQMRALPLLDLPVVEVAQGSSS 1762
Qy 1592 GQ-----KVISNVGKGTDAANVQOLNEVNNLLGLGNAGNDADG---1633
Db 1763 NGDGNSTQSSSTTTTTTTTTSSDGGESTTSSDP--VVEVSQGTN-----CDNSSQSS 1812
Qy 1634 -----NOVNTADIK-----KDPNSGSSSNRTVIKAGTVLGGKGNNDTE 1671
Db 1813 SSTTTTNEVSLKDNRPKNWRTTKYSSRTIRIPNSGRKLSNSESSTSTTV---TSSSSS 1869
Qy 1672 KLATGGIQGVYDKDGNANGDLSNVVVKQDKGSKKALLATYNAAGQTNVLTNNPAEADR 1731
Db 1870 KPQTKYSWSSSSKKSNNGGKKNKYWKRTKKSRKNNNGSSSTIVGE-----ESSDS 1920
Qy 1732 INEQGIRFFHVNDGNOE-----PVVQGRNGIDSSASG-----KHSVAIGF 1771
Db 1921 LTDAGVDVTOGNGLNDGESSQSTVTSSLPVVDTSADYQNSSESLSTSTENTTKYSKSEK 1980
Qy 1772 QAKADGEAAVAIGROTQAGNOSIAGDNNQAATGDSIAGTGNVNVAGHSGAIGDPSTVK 1831
Db 1981 VPKSNQSSSISASKTKTKVTSTSTSPNVKSSKK-----TSN--SGK---SVKTSSTTI 2030
Qy 1832 ADNSYSVGNNOFTDA--TQTDVFGVGNNTVYESNSVALGNSAISAGTHAGTQAKKSDG 1890
Db 2031 TTTSDPGOSSITOGIPONDIKSL--NQVTTTTSSVQGVPPSPVVKVTKETSVDKQK 2089
Qy 1891 TAGTTTTAGATCTVKGf--AGQTAAGVAVSGVAGSAGERRIQNVAAG-----EVSAT 1938
Db 2090 TTRSSTTTTTTTKGSNOGSLTLTLPADV--GLKSTKTTTTTKGTLKSLDILSLPEVDAS 2148
Qy 1939 STDVANGSOLYKATOSIANATNELDHRHQNEN--KANAGISSAMAMASMPQAYIPGRM 1996
Db 2149 I--AVNGDESRSASIKDTNILSKIDLPLKLDASLVNNGKSSKSSSTTTTTTTSTTKGNKV 2206
Qy 1997 -----VTGGIATHNGQCAVAVGLSK---LSDNGQWVFKINGSADTQK 2035
Db 2207 SLSLPEVDASTAV--NGDDARSASIKDTNILSKIDLPLKLDASLVNNG 2253
RESULT 14
Q55582
ID Q55582 PRELIMINARY; PRT: 3029 AA.

Thu Sep 13 14:19:01 2001

AC Q55882; 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 308.8 KDA PROTEIN.
GN SLR0364.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA Tabata S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
Sugliura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
Miyajima N., Hirose M., Sugliura M., Sasamoto S., Kimura T.,
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D63999; BAA10087.1; --
KW Hypothetical protein.
SQ SEQUENCE 3029 AA; 308789 MW; BE0BB57E822A4FB CRC64;

Query Match 5.08; Score 518; DB 2; Length 3029;
Best Local Similarity 21.88; Pred. No. 2e-13;
Matches 521; Conservative 256; Mismatches 901; Indels 710; Gaps 119;

QY 56 VTGATL-----SGSAYA-----QKDKTHIAIGE-----QKQPR 84
DB 627 VSGATISVSGSVYTVTVTGIANNGVNLDFASGQNIIRDNTANNALSNTPTTDEQYT 686
QY 85 RSGTAKAGD---DRAIAGENAN-AQGQATAI-----GSSN--KTVNGSLDKIGTDAI 133
DB 687 LONTAPASITILDANITADDDIINIAESGOAIPITGVGGEFNVGDTVTLTVNDKTFAGV 746
QY 134 GQESI---AIGDDVKASDASIAIGSDDLHLLDGHGPKHPKGLINDLINGHAULKEI 189
DB 747 GAGGLESINVPGLDVLVDADLPIA---SIATDAAGN-----L 782
QY 190 RSKKONDVYRRTA-----SGHASTAVGAMS----- 216
DB 783 GSATDQNTVDTTAPITPIITVNDVTADNIIINAEASGOAIPITGVGGEFNVGDTVTLTV 842
QY 217 -----YAOGHES-----NAFGRATASAYSLAVGLAATAGOSTIAIGSDA 258
DB 843 NGPFTGTVDANGDFSIDVLGDLVNGSLDTIAASVAITDAAAGNPGSASDNGTYVDTTA 902
QY 259 TSSSLGAIALG-----ACTRAIQOG-SIALGQGSVVVTS--DNNSRP-----AYTNT 303
DB 903 PTVTINAVDDDIINAVEAGSPVAVSGTTGVEDGQVTVTIDGNTYATVGTGNATWNI 962
QY 304 QALD-PKFOATNTKAGPLSIGNSIKRKLIINVGACVKNKTDVAVNAQLEAVVVKWAKERRI 362
DB 963 PVADIANFEATEEVAVPVSIDLGNPATRNII-----IVDIVAPA-----V 1004
QY 363 TFGDDNSTDVK-----IGLDNTLTIKGAETNALTDNNIIGVVKKEADNSGLVKLAKTLNN 418

Db 1005 TIDSISDDTGAQANDFTINDDTLVFNGTAEADST-----VVVSLDG----- 1045
QY 419 LTEVNTTTLNATTVTKVCSSSST---TAEILLSP-----SLTETOPNTGSQSTSKTV--- 466
Db 1046 -IEIGTVTAN-----GAGEWLDTGTLLADGDVLSVTATNPTGNSATATQTIWVDI 1097
QY 467 ----YGVNGVKFTNNAETTAAGITRITRDKIGFARDGDVDEKQAPYLDKKKQLKQSVVAI 522
Db 1098 TAPVTINAVDDDIINAVEAGSPVAVSGTTG-VEGQV-----VTV 1139
QY 523 TID-----NGIDAGNKKISNLAKSSANDAVTIEOLK-----AAKPTLNAGAGISVPT-- 571
Db 1140 TIDGNTYATVGTGNATFNIPVADIANFEATEEVAVTSDLAGNPATPATRNITVDTTAP 1199
QY 572 -----EISVD-----AKSGNVT-----APTNYIGVKTTELNSDGTSDKFSVK 608
Db 1200 TVTINAVDDDIINAVEAGSPVAVSGTTGVEDGQVTVTIDGNTYATVGTGNATWNI 1259
QY 609 GSGTNNSLVTAHILASYLNEVNRRTADSALQSFTVKEEDDDANATV-----AKD 658
Db 1260 VADIANFEATEEVAVTSDLAGNPATPATRNITV-----DTVAPAVTIDSISDDTGAQAND 1315
QY 659 TTKNAGAVSILKLGKGLTVATKKGDTVTEGLSQDSGLTIGKSTLNNDGLTVKDTNEQI 718
Db 1316 FITNDDTLVF-----NGTAEA---DSIVVYSL---DGEIGTVTANGAGEWTLDTYGT 1363
QY 719 QVGANGIKFTVNGSNPGTGIANTRITRDKIGFAGSDGAVDTNKPYLDDQKLOVGNVKI 778
Db 1364 L--ADGDVLSVTATNPG--NSATATQTIV-----VDTTAPTVTINAVDDI-- 1408
QY 779 TWYGINA--GGKAITGLSPTLPSIADQSSRIEL--GNTIQDKDKSNAASIN----DILNT 831
Db 1409 ----INAVEAGSPV--AVSGTTGVEDGQVTVTIDGNTYATVGTGNATWNIPIVADIAN- 1462
QY 832 GFNLKNNNPIDFVSTYDIV----DFANGNATAT--VTHDTANKTSKVVDVWVDD--- 882
Db 1463 -----FEATEEVAVTSDLAGNPATPATRNITVDTTAPT-VTINAVAVDDIIN 1509
QY 883 -----TTIHLTGTDONKGLGKVTTKLNTKTSANGTAT-----NFNVNSD----- 922
Db 1510 AVEAGSPVAVSGT---TTGVEDGQVTVTIDGNTYATVGTGNATWNIPIVADIANFEAT 1565
QY 923 EDALVNAKDIAENLNTLAKEIHTTKGTADTALOTFTVKVVDENNADNANATVQKNNAN 982
Db 1566 EEVAVTSDLAGNPATPA---TRNITVTVAPAVTIDSISDDTGAQANDFIT----- 1614
QY 983 NQVNTLTLKGELNLTCKDKNGTVTEGINTTSLKAGKSTLNDGGLSINKNPTGSEQIOVG 1042
Db 1615 -NDDTLVFNQ-----TAEADSTVVVSLD---GIEIGTVTANGAGEWTLDTYGT---LL 1660
QY 1043 ADGVKFAKVNNGVWAGIDGTRITRDEIGFTGTNGSLDKSKPHLSKDGINA--CGKKI 1100
Db 1661 ADGDVLSVTATNPTGNSATATQTIWVDITAPTVTINAI-----AVDDIINAVEAGSPV 1714
QY 1101 -----TNIOSEIAQ-----NSHDAVTGKIYDLKTELEN--KISSTAKTAQNSLHFS 1147
Db 1715 AVSGTTGVEDGQVTVTIDGNTYATVGTGNATWNIPIVADIANFEATEEVVA-----T 1768
QY 1148 VADSGN-----NETVSN--PYSSYDTSKTSQVITF---AGE-----NGITTKVNGK-VV 1191
Db 1769 VSDLAGNPATPATRNITVDTTAPTVTINAVDDDIINAVEAGSPVAVSGTTGVEDGQV 1828
QY 1192 RVGIDQTKGLTTPKLTPLVGNNGKGI-VIDSQNGNTITGLSNTLANVTNDKGSVRTTEOG 1250
Db 1829 TVTID---GNTYATVGTGNATFNIPVADIANFEAT-----EEVAVTSDLAGNPATPATR 1881
QY 1251 NI KDEKTRASTVDVLSAGFNLOGGEAVDFVSTYDVTNFDAGNA---TTAKVTYDDT 1307
Db 1882 NI--TVDTVAPAVTIDSISDDTGAQAN-----DFITNDTTLVF--NGTAEADSTVVVSLDGI 1934
QY 1308 ---SKTSKVVDVWVDDTTIEVKDKKLGKVTTLTTLTGANKFALSNOATGDALYKASD 1364

QY 646 DDDANAIVAKDTTKN--AGAVSILKLGKNGLTVA---TKKGTVTTFG--LSODSOLT 698
 Db 725 DIDNSGLVAGELENTLSDGSLVK-TGTGELTSGDNTYSGGTTITGTTLADHADS 783
 QY 699 IGKSTLNDGL-----TVKDTNEIOVANGIKFTNVNGSNPGTGTANTARTRD 748
 Db 784 LGSGLDINSGLVKVGBEDLENTLSGSGSLVKGTGTGELTSGGNDYSGGTTIIG-CTLTAD 842
 QY 749 KICFAGSGAVDTNKPYLDQDKLQGVNKTITNTGNAGKAITGLSP-TLPSIADQSRN 807
 Db 843 HADSLGT-GAV-----ANSGLVQGELENTLSGSGSLVKGTGTGELTSGDNTYSGGT 895
 QY 808 IELGNTIQDKKSNASINDILNTGNLKNNNPDI FVSTYDIVDFANGNATTAIVTHTD 867
 Db 896 TIIGGTL-----TADHADSLGT-----AVANSGLVQGELENTLSGSGS 937
 QY 868 ANKTSKVVYDVNVDTTIHLTGDDNKKLGKVTTKLNKT SANGNTATNFVNSDEDALV 927
 Db 938 LVKTG-----TGELTSGDNTYSG-----GTTIIGTLTADHADSLGTGAVA 979
 QY 928 NAKDI-----AENLNTLAKEHTTK-GTADTAL---OTFTVKKVDENNNADANAITVGOK 979
 Db 980 NSGLVQGELENTLSGSGSLVKGTGTGELTSGDNTYSGGTTITDGLIAANVNALGSG 1039
 QY 980 NANNQVNTLTKGNG---LNKTKNGTVTFGINNTSLKAGKSTLNDGG-LSIKNPT 1034
 Db 1040 NIDNS-CTLILDA-NGAFELANVTTHSGATTALAGST--LDAGQFTQEDGSLTSLDLGA 1095
 QY 1035 GSEIOVAGDGVKPAKVNNGVVGAGIDGTRITRDEIGFT--GTNGSLDKSPHLSKDG 1092
 Db 1096 ATDDAVITADSVALGGTLN-VTGIG-SVTDSTWTPAYTYTLIGSDSATTDFDGLTVAG 1152
 QY 1093 INAGKKITNTQSGEI--AONSHDPAVTGKLYDLKTELENKISSTAKTAQNSLHEFSVAD 1150
 Db 1153 MNREDDVFLTI-DGKVDADNTH-----YDL-----TASLWYADRDNATTD 1193
 QY 1151 EOGNNFTVSNPYSDYDSTKT-SDV-----ITFAGENGIT--TK 1185
 Db 1194 AHG-TFTLSDPGSGFNVAATLTDVDDTLDPGSRWDGKSLTKEGAGTLILSGDNDYSGGTT 1252
 QY 1186 VNKGVVRV-----GIDOTKGLTTPKLTGVNNGKG- 1215
 Db 1253 INEGLTVAASTALGTGLVDNNATLVLADAGEVSAVGITHTHGATT-QLALGTSLDLGD 1311
 QY 1216 -----IVIDSONGNTITGLSNTLAN---VTNDKGSVRTTE-QCNIILKDEKDT 1259
 Db 1312 SALIOQDGSSTLVELNLSDSVQPLITGSSATLGGDLVVSASLQARASDAEFQSEKLMDMT 1371
 QY 1260 RAASIVDVLISAGFNLOGGEAVDFVSTYDVTNFDAG-----NATKAVTYDDTS 1308
 Db 1372 SDIS-GDFTSLTMNLT---DKPDYLTVTCTINPEDASEYLLTEGLSWNATATSAT----- 1422
 QY 1309 KTSKVYDVNVDTTIEVKDKKLGKVT-----TTLTSTGTGANKFALSNOATGDALVK 1361
 Db 1423 -PAHGTFTLAGD-SFEV-TSVLGDKTGNGDWDGKSLTKLGAKLTLSGANTYSGDINQV 1479
 QY 1362 ASDIHAHLMTLSGDIQTAK-CASQANNAGYVDADGNKVIYDSTDNKYVQAKNDGTVDKT 1420
 Db 1480 EGTG-----WLSGDSGIGBMGQQAVNVASGATFG-----SNGTTVNG 1518
 QY 1421 KEVADKLVQAQATPDGTLAOMNVKS-----VINKEOVNDANKKOGINE---DNAFVKG 1471
 Db 1519 K-----VTNEGTLVFNSEETGAFTFNGDLINMGTMWTSGSSSTPGNTLYVDG 1567
 QY 1472 LEKAASDNKTKNAA-----VTVGDNLNAQAQPLTFAGD-TGTTAKKLGETLTIKG---GQ 1522
 Db 1568 -----NYTGNGGSLYLVGLDSDS-ATDKLVITGDASGTT-----DLYINGIGDGA 1613
 QY 1523 TDTNKLTDNNIGVWAGTDCFTVKLAKDLTNLNSVAG-----GTFKIDDKGVS 1569
 Db 1614 QTTNGIEVYDVGVSDFELK-----NEVNAGLYTYRLYNESDNDWYLASKAQS 1665

QY 1570 FVDSSGOAKANTPVLISANGLDLGGKVISNKGKTKDTKDTAANVOOLNEVRNLLGLGNAGN- 1628
 Db 1566 DDDDSG--GDDTp---SDGGDDGGNV-----TPPDDG-----GDGGNV 1698
 QY 1629 ----DNADGNOVNITADIKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLATGGIQGVVDK 1684
 Db 1699 TPPDDGGGDDVTPPD-----HGGDVAPOYRADITGAYMGWQWARNLQM-----QTLYDR 1748
 QY 1685 DG-----NANGDLSNVVVKTKQDGSKKALLATYNAAGOTNY-----LTNNPAPAID 1730
 Db 1749 EGSQYRNADG---SVWAFK-KAKAESEAVSGNIDMDSNYSQFOLGGDILAWNGOOSV- 1803
 QY 1731 RINBOGIRFFHVNDGNOEPVVGQRNGRIDSSASGK-HSVAIGFQAKADGEAAVAIGBQTA 1789
 Db 1804 ---TVGVMASVINADTDSGTNGRGADGSGQFTSSGNVDGYNLGVYATWFPADA-----QTHS 1854
 QY 1790 G-----NOSTAIGDGAQATGQDSTAICTGNVWAGKHSAGTGPSTVKADNSYS 1837
 Db 1855 GAYVDSWYQYGFYNNVESGDAGSEYD-----STANNV-----SLETGYRYD 1897
 QY 1838 VGNNOFTDATQTDVFGVGNNTVYTESNSVALGNSAISAGTHAGTQAKKSDGTAGTT 1895
 Db 1898 IALSN-----GNTVSLTPQAQVVMQNYSDSVKDNYGTRIDGQDGSWTT 1942

Search completed: September 13, 2001, 12:55:10
 Job time: 760 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 22:50:01 ; Search time 33595.7 Seconds
(without alignments)
2835.657 Million cell updates/sec

Title: US-09-361-619-8
Perfect score: 6159
Sequence: 1 atgaatcacatataaagt.....ttggtgcagggttttcacttt 6159

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_vil2.*
60: gb_vil3.*
61: gb_vil4.*
62: gb_vil5.*
63: gb_vil6.*
64: gb_vil7.*
65: gb_vil8.*
66: gb_vil9.*
67: gb_vil10.*
68: gb_vil11.*
69: gb_vil12.*
70: gb_vil13.*
71: gb_vil14.*
72: gb_vil15.*
73: gb_vil16.*
74: gb_vil17.*
75: gb_vil18.*
76: gb_vil19.*
77: gb_vil20.*
78: gb_vil21.*
79: gb_vil22.*
80: gb_vil23.*
81: gb_vil24.*
82: gb_vil25.*
83: gb_vil26.*
84: gb_vil27.*
85: gb_vil28.*
86: gb_vil29.*
87: gb_vil30.*
88: gb_vil31.*
89: gb_vil32.*
90: gb_vil33.*
91: gb_vil34.*
92: gb_vil35.*
93: gb_vil36.*
94: gb_vil37.*
95: gb_vil38.*
96: gb_vil39.*
97: gb_vil40.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6159	100.0	6159	10	AX079920 Sequence
2	4101.6	66.6	6141	10	AX079918 Sequence
3	4101.6	66.6	6972	10	AX079917 Sequence
4	4090.6	66.4	6973	9	AR040716 Sequence
5	4090.6	66.4	6973	10	AX079913 Sequence
6	3982	64.7	6209	10	AX067457 Sequence
7	3947.8	64.1	5976	10	AX079914 Sequence
8	3946.8	64.1	6259	10	AX079924 Sequence

Db	3301	GGCAATTAACGCGAGGTGGTAAAAAGATTACCAACATTCATCAGGTGAGATTGCCAAAAC	3360
Qy	3361	agccatgatctgtgacagcgccgaagatttatgatatttataaaacccgaacttgaaataaa	3420
Db	3361	AGCCATGATGCTGTGACAGCGCGCAGATTATGATTTTAAANACCGAATTTGMAATAAA	3420
Qy	3421	atcagcagtlactgcccacacagacacaaactcattacacgaattctcagtagcagatgaa	3480
Db	3421	ATCAGCAGTACTGCCAAAACACACAAAACATCATTTACACGAATTTCTCAGTAGCAGATGAA	3480
Qy	3481	caaggtaataactttccggttagtaacccttactccagttatgacacctcaagacctct	3540
Db	3481	CAAGGTAATAACTTTTACGGTTAGTAGTACCCTTACTCCAGTTATGACACCTCAAGACCTCT	3540
Qy	3541	gatgctcaccttttcagtgtaaacgagcattaccacacacaggttaataaaagtggtg	3600
Db	3541	GATGTCATACCTTTTCAGGTTGAANACGGCATTTACCACCAAGTAAATAAAGGTGTGGTG	3600
Qy	3601	cgltggtggtcattgac	3660
Db	3601	CGTGTGGGCATTGACCAAAACCAAGGCTTAACACACCCCTAAGCTGACCGTGGTGAATAAT	3660
Qy	3661	aatgcaaaagcattgtcattacacagcaaaatggtcaaaataccatcacaggaactaagc	3720
Db	3661	AATGGCAAGGCATTGTCATTACAGCCAAATGGTCMAAATACCATCACAGGACTAAGC	3720
Qy	3721	aacactctagtaattgtaac	3780
Db	3721	AACACTCTAGCTAATGTTACCAATGATAAAGGTAGGTAGCGCACCCACAGAACAGGCAAT	3780
Qy	3781	ataatcaaaagcagaac	3840
Db	3781	ATAATCAAAAGCAGAAC	3840
Qy	3841	tttaacttgcaagcaaatggtgaagcggttgactttgtctccacttatgacacacacacac	3900
Db	3841	TTTAACTTGCAAGCAATGGTGAAGCGGTGACTTTGTCTCCACTTATGACACCGTCAAC	3900
Qy	3901	tttgccaatgccaatac	3960
Db	3901	TTTGCCAATGGCAATGGCAATGGCAATGGCAATGGCAATGGCAATGGCAATGGCAATGGCA	3960
Qy	3961	aaagtggctctatgactgcaatgtagatatacaacacacacacacacacacacacacacac	4020
Db	3961	AAAGTGGTCTATGATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG	4020
Qy	4021	ggcgtataaaac	4080
Db	4021	GGCGTAAANACACACACATTTGACAGTACTGGCAGAGGTGCTATTAATTTGCCCTAAGC	4080
Qy	4081	aatcaagctactggcgatgctgcttgcagcgccagtgatctggtgctcatctataacacac	4140
Db	4081	ATNCAAGCTACTGGCGATGGCGCTTGTCAAGGCGAGTGCATGATGTCATGTCATGTCATG	4140
Qy	4141	tatatctggcgacatccaaactgcacacacacacacacacacacacacacacacacacacac	4200
Db	4141	TTATCTGGCGGACATCCAACTGCCAAAGGGCAAGCAAGCGAACTCACTACAGGCTAT	4200
Qy	4201	gtgagctgctgagtggaataaagtcactatgacagtlaccgatacaaaagtaactataagcc	4260
Db	4201	GTGGATGCTGTAGTGGCANTTAGGTCATCTATGACAGTACCAGTAACTACTATCAAGCC	4260
Qy	4261	aaaaatgatggcacagtgtgataaaacacacacacacacacacacacacacacacacacacac	4320
Db	4261	AAAAATGATGGCACAGTTGATAAAACCAAGAGGTGGCCAAAGACAACTGGTCGCCCAA	4320
Qy	4321	gcccaaac	4380
Db	4321	GCCCAAAACCCAGATGGCAGATTTGGTCAAAATGAATGTCATCAATCAGTCATTAACAAGAA	4380
Qy	4381	caagtaaatgatgccataaaagcagaagcagcatcaaatgaagacacacacacacacacacacac	4440
Db	4381	CAAGTAAATGATGCCAATAAAAGCAAGGCATCAATGAAGACACACGCTTTCTTTAAGGA	4440
Qy	4441	cttgaaaaagccctctctatacaaaaaacacacacacacacacacacacacacacacacacacac	4500
Db	4441	CTTGAAAAAGCCCTCTCTGTATAACAAAAACCAAAACGCCGAGTAACCTGTGGGTGATTTA	4500
Qy	4501	aatgcccgttgcac	4560
Db	4501	AATGCCGTTGCCCAACACCGCTGACCTTTGTCAGGGGATACAGGCACACGCTTAAAAAA	4560
Qy	4561	ctggcgagacttttgaccatcaaaagtggaacacacacacacacacacacacacacacacacacac	4620
Db	4561	CTGGCGAGACTTTTGACCATCAAAAGGTGGCAACACACACACCAATTAAGCTAACCGGATAAT	4620
Qy	4621	aacatcggtggtgtagcaggtactgattggtctcactgtcaaaacttgcacaaagccctaaccc	4680
Db	4621	AACATCGGTGTGGTAGCAGTACTGTGCTTAACACTTGCCTCAAACTTGCCTCAAAAGACCTAAC	4680
Qy	4681	aatcttaacagcgttaatgaggtggtgacacacacacacacacacacacacacacacacacacac	4740
Db	4681	AATCTTAACAGCGTTAATGTCAGGTGGCACCACCAAAATGATCAAAAGGCATCTCTTTTGTGA	4740
Qy	4741	gacgcaaacggtcaagcacaac	4800
Db	4741	GACGCAAAAGGTCAAGCCAAAGCAACACACCCCTGTGCTTAAGTGCCAATGGCTGGACCTG	4800
Qy	4801	ggtggcaaggttcaatgaggtggtgac	4860
Db	4801	GGTGCAAGGTTCATCAGTAATGTCGGCAAGGCACAAAGATACCGACGCTGCCAATGTGA	4860
Qy	4861	caacagttaaacgaggtacgcaactgttggtcttggtatgataacgctgacggtgacacacacac	4920
Db	4861	CAACAGTTAAACGAAGTACGCAACTTGTGGGTCTTGGTAAATGATACGCTGACGGCAAT	4920
Qy	4921	caggttaacacattgcccagac	4980
Db	4921	CAGGTAACATTGGCGACATCAAAAAGACCCCAATTCAGGTTTCATCATCAACCGCACT	4980
Qy	4981	gtcatcaaaagcagcaggttactgtggtggtggtggtggtggtggtggtggtggtggtggtggt	5040
Db	4981	GTTCATCAAAAGCAGGACGCTACTTGGCGGTAAAGGTAAATACGATACCGCAAAACTTGGC	5040
Qy	5041	actgggtgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	5100
Db	5041	ACTGGTGGTGTACAAAGTGGCGCTGGATAAAGACGGCAACGCTTAACGGCGATTTAAAGCAAT	5100
Qy	5101	gttgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	5160
Db	5101	GTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTGCCACTTTATAACGCC	5160
Qy	5161	gaaggtcagac	5220
Db	5161	GCAGGTGACAGCAACATATGTGACCAACACCCCGCAGAAAGCCATTTGACAGAAATAATGAA	5220
Qy	5221	caaggtatccgctcttccatgacaggtgacacacacacacacacacacacacacacacacacacac	5280
Db	5221	CAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCCTGTGTACAAGGCGT	5280
Qy	5281	aacggaatgaactcaagtcgctcagcgaagcactcagtggtggtggtggtggtggtggtggtggtggt	5340
Db	5281	AACGGCATTCAGTCAAGTCCCTCAGGCAAGCACTCAGTGGCGATAGGTTTCCAGGCCAAG	5340
Qy	5341	gcagatggtgaaagccgctgtgcatagcagacacacacacacacacacacacacacacacacacac	5400
Db	5341	GCAGATGTTAAAGCCGCGTGGCATAGGCAGACAAACCCCAAGCAGGCAACCAATCCATC	5400
Qy	5401	gcatcggtgaataacgcac	5460
Db	5401	GCCATCGGTGATACGCACAAAGCCACCGGCGATCAATCCATCCGCTCGGTACAGGCAAT	5460
Qy	5461	gtgtagcaggtgaagcactctgtgcatcggtggtggtggtggtggtggtggtggtggtggtggtggt	5520
Db	5461	GTGGTAGCAGGTAAAGCACCTGTGTGTCATCGGCACCAAGCACTGTTAAGGCTGATNAC	5520

Db 3154 AATAAATGTTGTTAGGTCCTGGGCTTATGATGGCAACATCGCATATACAGAGATGAA 3213
|||||
Qy 3241 attgctttactgggactaaTggctacttgaTaaagcaaaacccacccaaagcaagac 3300
|||||
Db 3214 ATTGGCTTTACTGGGACTAATGGCTCACTTCATATAAAGCAAAACCCACCTTAAGCAAGAC 3273
|||||
Qy 3301 ggcattaacgcaggtgtgtaaaagattaccacaattcaatcaggtgagattgcccataaac 3360
|||||
Db 3274 GGCATTAAACGCAGGTGTGTAAGATTACCAACATTCATCAAGGTGAGATTGGCCCAAAAC 3333
|||||
Qy 3361 agccatgatctgtgacagcgcaagatttatgatttaaaacccgaactgaaataaa 3420
|||||
Db 3334 AGCCATGATGCTGTGACAGCGGCAAGATTATGATTATAAACCCGAACCTTGAAAAACAA 3393
|||||
Qy 3421 atcagcagttactgcacaaacagcacaaaactcattcacgaattctcagtagcagatgaa 3480
|||||
Db 3394 ATCAGCAGTACTGCCAAACAGCACAAACACTATTACACGAATTCCTCAGTAGCAGATGAA 3453
|||||
Qy 3481 caaggttaactttagcgttagtaacccttactccagttatgacacctcaaaagacctct 3540
|||||
Db 3454 CAAGGTAAATACCTTTACGGTTAGTAACCTTACTCCAGTTATGACACCTCAAAAGACCTCT 3513
|||||
Qy 3541 gatgtatcccttgcaggtgaaacgcgattacacaaaggttaataaaggtgtggtg 3600
|||||
Db 3514 GATGTCAATCACTTTTCAGGTGAAACGGCATTTACCAACCAAGGTAAATAAGGTGTGGTG 3573
|||||
Qy 3601 cgtgtggcattgacacaaacaaaggtctaacacacgcttaagctgacgctgggttaataat 3660
|||||
Db 3574 CGTGTGGCATTTACCAAAACCAAGGCTTAACCAACGCTTAAGCTGACCGTGGGTAAATAAT 3633
|||||
Qy 3661 aatggcaaggtcattgtattaaacagcgaataatggtcgaataatcattcacagggactaagc 3720
|||||
Db 3634 AATGGCAAGGCATTTGTCATTGACAGCCAAATGCTCAAAATACCATCACAGGACTAAGC 3693
|||||
Qy 3721 aacacttagctaatgttaccatgataaaggtagctacgacacagacagaacagggcaat 3780
|||||
Db 3694 AACACTCTAGCTAATGTTTACCATGATAAAGGTAGCTAGCTACGCACACAGAACAGGGCAAT 3753
|||||
Qy 3781 ataataaagacgaagacaaacccgtgcccagcagcttgttgatgtctaaagcgacggc 3840
|||||
Db 3754 ATAATCAAGACGAAGACAAACCCGTCGCCAGCATTTGTGATGTGCTTAAGCGCAGGC 3813
|||||
Qy 3841 tttaactgcaaggaatgtgtgaagcgttgactttgtctccacttatgacacogtcaac 3900
|||||
Db 3814 TTTAACTTGAAGCAATGTGTGAAGCGGTGACTTTGTCTCCACTTATGACACCGTCAAC 3873
|||||
Qy 3901 ttgccaatggcaataccaccacgcgttaagtgacctatgatgacacagcaaaacacagt 3960
|||||
Db 3874 TTTGCCATGGCAATGCCACCCGCTAAGGTGACCTTATGATGACACAAAGCAAAACACAGT 3933
|||||
Qy 3961 aaagtgtctatgatgtcaatgtgatgatacaaccattgaagttaaagataaaaaactt 4020
|||||
Db 3934 AAAGTGTCTATGATGTCAATGTGATGATACACCATTTGAAGTTAAAGATAAAAACTT 3993
|||||
Qy 4021 ggcataaaacccacacattgaccagttactgacagcaggtgtcgaataatttgcacctaaagc 4080
|||||
Db 3994 GGCCTAAAAACCCACATGACACGACTGACGACAGGTGCTTAATAATTTGGCCCTAAGC 4053
|||||
Qy 4081 aatcaagctactggcagtcgcttgcgaagcgaagtcagttatcgtgtctatcttaaacacc 4140
|||||
Db 4054 AATCAAGCTACTGGCGATGGCGCTTGTCAAGGCCAGTATATGCTGTGCTATCTATAAACACC 4113
|||||
Qy 4141 ttatctggcgacatcccaactgccaaaggggcaagcgaagcgaacaaactcagcagggctat 4200
|||||
Db 4114 TTATCTGGCGACATCCAAACTGCCAAAGGGGCAAGCCAAACGGAACAACACTCAGCAGGCTAT 4173
|||||
Qy 4201 gtgagtctgatggcaataaggtcatctatgacagttaccgataaacaagtaactatcaagcc 4260
|||||
Db 4174 GTGGATGCTGATGGCAATAAGGTCTATCTATGACAGTACCAGTATACCAAGTACTATCAAGCC 4233
|||||
Qy 4261 aaaaatgatggcacagttgataaaaccaaagaagtttcccaaaagacaaactggtcgcccaa 4320
|||||

Db 4234 AAAAATGATGCGACAGTTTGATAAAACCAAGAAAGTTGCCAAAGACAAACTGGTGGCCCAA 4293
|||||
Qy 4321 gccaaaccccgatggcacattgctcaaatgaatgtcaaatcagttcatlaacaaagaa 4380
|||||
Db 4294 GCCAAACCCCGATGGCATTGGCTCAATGAATGAATGTCAATCAGTCATTAAACAAGAA 4353
|||||
Qy 4381 caagttaaatgatgccataaaaaagcaaggcatcaatgaagcaaacgcctttgtttaagga 4440
|||||
Db 4354 CAAGTAATGATGCAATAAAAGACAGGCATCAATGAAGACAAACGCCCTTTGTTAAAGGA 4413
|||||
Qy 4441 cttaaaaaagccgcttctgtatacaaaaaaccccaaacccgcagtaaacgtgtgggtattta 4500
|||||
Db 4414 CTTGAAAAAGCCGCTTCTGTATAACAAAACCAACCGCCAGTAACTGTGGGTGATTTA 4473
|||||
Qy 4501 aatccgcttgcccaaacacccgctgacctttgcaggggatacaggcacacgcgtlaaaaaa 4560
|||||
Db 4474 AATCCGCTTGCCCAACACCGCTGACCTTTGCAGGGGATACAGGCACACGCTTAANAAA 4533
|||||
Qy 4561 ctggggagacatttgaccatcaaaaggtgggcaaacagacacccaataaagtaaacogataat 4620
|||||
Db 4534 CTGGCGAGACTTTGACCATCAAAAGGTGGCAACACAGACACCAATAAGCTAACCGATAAT 4593
|||||
Qy 4621 aacatcggtgtgtgagcaggttactgatgcttactcgtcaaacitggccaaagacctaac 4680
|||||
Db 4594 AACATCGGTGTGGTAGCAGGTACTGTAGGTTCACCTGTCAAACTTTGCCAAAGACCTTAACC 4653
|||||
Qy 4681 aatcttaacagcgttaattgcaggttgccacaaaattgtatgaagaagggcatctctttgta 4740
|||||
Db 4654 AATCTTAACAGCGTTAATGCAAGGTGGCACCAAAATTTGATGACAAAGGCGTGTCTTTTGTGA 4713
|||||
Qy 4741 gacgcaaacggtcgaagccaaagcaaacacccctgtgctaagtgcgaattgggctggacgtg 4800
|||||
Db 4714 GACTCAAGCGGTCAAGCCAAAAGCAACACCCCTGTGCTTAAGTGCCTAATGGCTGGACCTG 4773
|||||
Qy 4801 ggtggcaaggttcacagtaattgtgggcaaaagccaaaagataccgacgcgtgccaatgta 4860
|||||
Db 4774 GGTGGCAAGGTCTACAGTAATGTGGGCAAGGCAACAAAGATACCGACGCTGCCAATGTA 4833
|||||
Qy 4861 caacagttaaacaggttacgcaactgttgggtctt-----tggttaatgataacgct 4911
|||||
Db 4834 CAACAGTTAAACGAAGTACGCAACTTGTGGTCTTGGTAAATGTGGTTAATGATFANACGCT 4893
|||||
Qy 4912 gacggcaatcaggttaaacattgcccagacatcaaaaaagaccccaaatctcaggtttcatctct 4971
|||||
Db 4894 GACGGCAATCAGGTAAACATTCGCGACATCAAAAAGACCCCAAAATTCAGTTTCATCATCT 4953
|||||
Qy 4972 aaccgacgttcatacaagcagcagcagctacttggcgttaaggttaaacgataaccgaa 5031
|||||
Db 4954 AACCGCACTGTCTATCAAAAGCAGCAGCGTACTTTGGCGGTAAAGGTAATAACGATACCCGAA 5013
|||||
Qy 5032 aaacttgccactgggtgtgtacaagtggtgataaagcgcgcaacgcgtlaacggcgat 5091
|||||
Db 5014 AAAGTGGCACTGGTGTATACAGTGGCGGTGGATTAAGACGCGCAACGCTTAACGCGCAT 5073
|||||
Qy 5092 ttaagcaatgttgggttcaaaaaaccccaaaagatggcagcaaaaaagccctgtctgcact 5151
|||||
Db 5074 TTAAGCAATGTTGGGTCAAAACCCCAAAAGATGCGACGCAAAAAGCCCTGCTCGCCACT 5133
|||||
Qy 5152 tataacgcgaggtcagacacaaactatgtgaccaacaaaccccgagaagccattgacaga 5211
|||||
Db 5134 TATAACCGCGCAGGTCTAGACCAACTATTTTGACCAACAAACCCCGCAGAGCCATTGACAGA 5193
|||||
Qy 5212 ataaatgaacaaggtatccgctcttccatgtcaacgtatggcaatcaagagcgtgtgta 5271
|||||
Db 5194 ATAATGAACAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCCTGTGGTA 5253
|||||
Qy 5272 caagggcgtaacggcattgactcaagtgcctcaggaagcagcactcagttggcgtataggtttc 5331
|||||
Db 5254 CAAGGGCGTAACGGCTTGAAGTCAAGTGCCTCAGGCAAGCACTCAGTGGCGCATAGGTTTC 5313
|||||
Qy 5332 caggccaaaggcagatggtgaagcgcgcgttgcctataggcagacaaaccccaagcaggaac 5391
|||||
Db 5314 CAGGCCAAGGCAGATGGTGAAGCCGCCCTTGCCATAGGCAGACAAACCCAGCAGGCAAC 5373
|||||

/organism="Moraxella catarrhalis"
/db_xref="taxon:480"

BASE COUNT	2265 a	1555 c	1532 g	1620 t	
ORIGIN					
Query Match	66.6%; Score 4101.6; DB 10; Length 6972;				
Best Local Similarity	81.3%; Pred. NO. 0;				
Matches	5063; Conservative 0; Mismatches 1009; Indels 156; Gaps 20;				
Qy	1	atgaatcacatcatataaagtcatttaacaaagacacagccacatttatggcgcg99cg 60			
Db	542	ATGAATCACATCTATAAAGTCATCTTTAAACAAAGCCACAGGCACATTTATGGCAGTGGCA 601			
Qy	61	gaatatcccaaatccccagctacgggggggtagctgtgctacagggcaagtggcagt 120			
Db	602	GAGTACGCCAATCCACACAGCAGGGGGGGTAGCTGTGCTACAGGGCAAGTTGGCAGT 661			
Qy	121	gtacgcactctaaagctttgcccgtattggcggcgctgcgtcgtctcctcgtgatcgagcg 180			
Db	662	GTATGCACTCTAGCTTTGCCCGCTATTGCCGCGCTGCTGTCTCGTGATCGGGTGCANCG 721			
Qy	181	ctcaatggcagctcttatgctcaacaaattactacca---agatcgaatttggcacaaca 237			
Db	722	CTCAGTGGCAGTGTATTATGCTCAAAAAGATACCAACATATCGCAATTGGTGAACAA 781			
Qy	238	aacaagataaaacaacgctgaaaggcgatgcccctagcgacaggtgaagcatccattgct 297			
Db	782	AACGACCCAAAGACGCTC-----AGGCACCTGCCAAGGCGGAGGTGATCGAGCCCTTTGCT 835			
Qy	298	tttgtagtcttcttaagcgacagggctctcaagctattgctatcggttagtgtaacacca 357			
Db	836	ATTGTTGAAATGCTTACGACACAGGGCGGTCAAGCCATCGCCATCGGTAGTAGTAATAA 895			
Qy	358	gatcctaataatggtagtaatggtaattgagtttccacgcccacaaaggtaacagatccatc 417			
Db	896	ACTGTCAATGGAAGCAGTTTGGATGAAGATAGTAGTACCGATGCTACGGGTCAAGAGTCCATC 955			
Qy	418	gcacatcggtggtagtatttgctgagggtagtgcctcgattgccatcggttagtgatgac 477			
Db	956	GCCATCGGGTGGTGTAAAGGCTAGTGTGCTCGATTCGCTCGATTCGATCGGTAGTGATGAC 1015			
Qy	478	ttatatattgcctaaagaatttggatctgaagaa-----tgaatttcacaaaa 522			
Db	1016	TTACATTTGCTTGTATCAGCATGTTGATTAAGTAACTCAACATCGCAAGAGTACTCTGATTAACGAT 1075			
Qy	523	cttattcatggccatgaaatattaaaaaaatatacaaacctcaacccatggttaaaatcaaa 582			
Db	1076	CTTATTTAACGGCCATGTCAGTATTAAAGAAATACCAAGCTCAAGAGGATAATGATCTAAAA 1135			
Qy	583	tatcgacgcacaagagacacaaggccacgcccagctactgagtggggagccatgcatatgca 642			
Db	1136	TATAGACGCACACCGCAAGCGGACACGCCAGTACTGTGCTAGTGGGAGCCATGTCATATGCA 1195			
Qy	643	cagggctattttccaaagcctttggtatcatcagcaacagctgaagctgcctattccttg 702			
Db	1196	CAGGGTCATTTTCCAAACGCTTTGTTACAGGGGCAACAGCTAANAAGTGCCTATTTCCTTG 1255			
Qy	703	gcagtaggtcttgcgcccacaaagccacaaacaaatcttcaatcgctgtgtgttcccaatgca 762			
Db	1256	GCACTGGGTCTTCCGCGCCACAGCCGAGGGCCAAATCTACAATCGCTATTGGTTCTGATGCA 1315			
Qy	763	aaagctaacgcgttttcgacgcagccattggtgggaaatactgtagtattttgggtcga 822			
Db	1316	ACATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGGAGGTACTCGTCTCAGCTACAGGGC 1375			
Qy	823	ggcgttgccttagttttgtctcagatccttgcagtaggagtaataatcacagatgcccagt 882			
Db	1376	AGTATTCCTCCCTAGGTCAAGGTTCTGTGTCTCAGTGTGATTAATAATTC---TAGACCG 1432			
Qy	883	gcctatgtaccactaggttaaaagcttagcagacaggtataaaagccaccgcagggatgat 942			
Db	1433	GCCTATACACCAAAATACCCAGGCCTAGAGCCCCCAAGTTTCAAGCCAC---CAATAATACG 1489			

Qy	5392	caatccatcgccatcggtgataacgcacaagccacggcgatcaatccatccatcgccatcggt 5451			
Db	5374	CANATCCATGCCATCGGTGATTAACGCACAAAGCCACGGGCGATCAATCCATCGCATCGT 5433			
Qy	5452	acagggcaatggttagcagtaagcaactgtgtgcacatcgccgaccccaagcactgtaag 5511			
Db	5434	ACAGGCAATGTGTAGCAGTAAGCACTCTGTTGCTCGCGACCGACCAAGCAGCTTTAAG 5493			
Qy	5512	gctgataacagtttacagttgggttaataacacacagtttaccgatgcccactcaaacagat 5571			
Db	5494	GCTGATAACAGTTTACAGTGTGGGTGATTAACAAACAGTTTACCGATGCCACTCAACCCGAT 5553			
Qy	5572	gtctttggttggtgcaataacatcacccgtgaccgaaagttaactcgtgtgcttaggttca 5631			
Db	5554	GTCTTTGGTGTGGCAATAACATCACCGTGACCGGAAAGTAACTCGGTTTGGCTTAGGTTCA 5613			
Qy	5632	aactctgccatcggtgagtcacacgcagcagcagcacaagcccaaaaaatctgacggcaca 5691			
Db	5614	AACTCTGCCATCAGTCAGGCACACAGCGAGGACACAAAGCCAAAANAATCTCAGCGCACA 5673			
Qy	5692	gcaggtacacacccacacagcaggtgcccacaggttacggtttaaaggctttgctggcaaacg 5751			
Db	5674	GCAGGTACACACCCACAGCAGGTCAACCGGTACGGTTAAAGGCTTTGCTGGACAAAGC 5733			
Qy	5752	gcggttggtggtgctccgtggtggtgctcaggtgctgaaacgcgtatcccaaatgtgca 5811			
Db	5734	CGCGTTGGTGGGCTCTCCGTGGGTGCTCAGGTGCTGAAGCGCTATCCAAAATGTGGCA 5793			
Qy	5812	gcaggtgaggtcaggtgcacacagcagcagtcggtgcaatggtgagcaggtgtacaaagcc 5871			
Db	5794	GCAGGTGAGGTGAGTGCACACAGCAGCAGTCCGTCATGATGATGAGCAGTTGTACAAAGCC 5853			
Qy	5872	acccaaagcatgcccacgcacaaatgagcttgacatcgtatccacaaacgaaat 5931			
Db	5854	ACCCAAAGCATTTGCCAGCGCAACCAATGAGCTTGACCATCGTATCCACCAACAAACGAAAT 5913			
Qy	5932	aaagccaatgaggtattcattcagcagtcggtgagtcggtccatgcccacaaagcctacatt 5991			
Db	5914	ANGGCCATGAGGGATTTTCATCAGCATGCGATGCGTCCATGCCACAAAGCCTACATTT 5973			
Qy	5992	cctggcagatccatggtttaccgggggtattgcccacccacaaagcagtcaggtgagtcga 6051			
Db	5974	CTGGCAGATCCATGGTTTACCGGGGGTATTGCCACCCACCAACGGTCAAGGTGCGGTGGCA 6033			
Qy	6052	gtggactgcgaagctgcgataatggtcaatgggttattttaaatacaatgggttcagcc 6111			
Db	6034	GTTGGAGCTGTCGAAGCTGTCGATTAATGCTCAATGGGTATTTAAATCAATGTTTCAGCC 6093			
Qy	6112	gatacccaaggccatgtagggggcgagttggtgaggttttcacttt 6159			
Db	6094	GATACCAAGGCCATGTAGGGGCGCAGTTGGTGCAGGTTTTCATCTTT 6141			

RESULT	3				
AX079917					
LOCUS	AX079917	6972 bp	DNA	PAT	22-FEB-2001
DEFINITION	Sequence 5 from Patent WO0107619.				
ACCESSION	AX079917				
VERSION	AX079917.1	GI:13159438			
KEYWORDS					
SOURCE	Moraxella catarrhalis.				
ORGANISM	Moraxella catarrhalis.				
	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;				
REFERENCE	1 (bases 1 to 6972)				
AUTHORS	Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.				
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella				
JOURNAL	Patent: WO 0107619-A 5 01-FEB-2001;				
FEATURES	CONNAUGHT LABORATORIES LIMITED (CA)				
	Location/Qualifiers				
source	1..6972				

Qy 943 tctacggatataatttttccatctggttaagtagtaataataatagcagtatcaggcgtaaa 1002
Db 1490 AAGCGGGTCCACTTTCCATGGTAGTAA-----CTCTATTCAACACGTAAA 1534
Qy 1003 atcatcaatgctggtgcgggtctcgggataccgataccggtcaatgtagcacacttaaa 1062
Db 1535 ATCATCAATGTCGGGTGCAGGTGTTAAATAAACCGATCGGTCAATGTGGCAGACGTAGA 1594
Qy 1063 ttggtggaggaaactgggcta-----atcgtaaaattacttttaagggtgtagtgaacaataa 1119
Db 1595 GCGGTGGTGAAGTGGCTAAGGAGCGTAGAATTACTTTTCAGGGTGATG-----ATAAC 1648
Qy 1120 agcaatagcgtagaaagaggttttgggcaactcttaactataaaaggtgtagcacagacc 1179
Db 1649 AGTACTACGCTAAATAAGGTTTGGATAAATACTTTAACTATTAAAGGTGGTGCAGAGACC 1708
Qy 1180 aacgattaaacggaactaaactcaggtggtgtaacagatgcaaa-----tggtctgaaa 1233
Db 1709 AACGATTAAACCGGATATATATCGGTGTGTAAGAGGCTGATATAGTGGTCTGANA 1768
Qy 1234 gttaaactgtgtaagagctgactggattgaccaggtgtctcc-----gct 1278
Db 1769 GTTAACTGTCTAAACTTTTAAACAATCTTACTGAGGTGAATACACTTAAATGCC 1828
Qy 1279 accaacaataatcccggttagtaataccaacaacaacacgagcgcgagctacaagcggtggt 1338
Db 1829 ACAACCCAGCTTAAGGTAGGTAGTAGTAGTACTACAGCTGAATATTAGGTGATAGT 1888
Qy 1339 ttgaccttttagccca-----ataacaggtacaaaaacagataaaaaacgctctacagcatt 1392
Db 1889 TTAACCTTTACCCAGCCCAATACAGGAGCTCAAGAGCAACAAACCGTCTATGGCGGTT 1948
Qy 1393 gatgattgaagtttactaataatagtagtaataatgatacaactaaaggctactactcgtatt 1452
Db 1949 AATGGGTGAAGTTTACTAATAATGCAGAAACACAGCAGCAATCGGCACCTACTCGTATT 2008
Qy 1453 accaaaaaagaataatggttttgcgtggttactaatgatggagttgatgaaagcaaaccttat 1512
Db 2009 ACCAGAGATAAATTTGGCTTTCGAGAGTGGTGTGATGTGTGATGATAAACAACAGCACCATPAT 2068
Qy 1513 cttaacacgaaaagctaaagattggcaacagacaccctaaacagtggttagcttaactgtt 1572
Db 2069 TTGATAAANAACAACTTAATAGTGGTAGTGTTCGAATTTACCATAGACAATGGCATTTGAT 2128
Qy 1573 aataacaccactggttaataaacaataatccagtcggtgctgataatgagcattaaatttggccaca 1632
Db 2129 GCAGGTATAAAGATCAGTAATTTGCCAAAGGTACAGTGTCTAACGATCGCGTTACC 2188
Qy 1633 gtcgctaataatgttgaaataacatcgcagcaacagtcggcactgctcgtattaccogaagag 1692
Db 2189 ATCGAACAGCTCAAGCGCCCAAGCCTACTTTAAACGAGCGGCTGGCATCAGTGTCTACA 2248
Qy 1693 aaaattggttttgcgtgtaactaatgatggagttgatgaacaagcaccataatttggataaa 1752
Db 2249 CCTACTGAATATCAGTTGATGCTTAAGAGTGGCAATGTATTACCGCCCCAACCTTACAACA-- 2306
Qy 1753 gaacgactaaagtggtgctgttgaaattaccacagatagtggtattaatgctggttaatt 1812
Db 2307 -----TTGGCGTGAAACACCCAGCGCTTAACAGTGATGCGCAGTGTGATAATTTAGT 2359
Qy 1813 cacaagattaccgacttactaataatggtatagcaataatcccgatgocggttaccatcaaacag 1872
Db 2360 GTTAAGGGTAGTGTACCAACAATAGCTTAGTTACCGCCGAACAATTTGGCAAGCTATCTA 2419
Qy 1873 ctcaaaagcgaagcctactttaaacgagcgatggcatcagattatagtaataaac 1932
Db 2420 AATGAAGTCAATCGAACGGCTGACAGTGTCTTACAAAGCTTTTACCGTTTAAGAGAAGAAC 2479
Qy 1933 ggggatctagttgatagtggtggaataattaccaccccaactataacattagcggtgaaa 1992
Db 2480 GATGATGAGC-----CCNACGGCTATCACCGTGCGT 2509

Qy 1993 accactaagcttaacagtaattggaaccagtggttaataataaatttagttagtaagtct 2052
Db 2510 AAAGATACGACAAATAATGCGCGCAGTCAGCATCTTTAAAACTCAAAAGTAAAAAACGGT 2569
Qy 2053 catgataacaatagcttagttaccgccaagatttggcagactatcctaataaagtcaat 2112
Db 2570 CTACGGTTGCTACCAAAAAGATGGTAGGTATTCTTTGGGCTTTAGCCAGATAGCGGT 2629
Qy 2113 gaaacggtgcagctgctctaccaagctttaaaagtcctcaaaacggttgataaagcaaac 2172
Db 2630 CTGACCATTTGCAAAAGCACCCTAAACAACGATGGCTTGTAAAGATACCAACGAA 2689
Qy 2173 gccatcacgctgggttaagataacaaacggaagacacctcaacaccttaaaactcaaaagt 2232
Db 2690 CAAATCCAAAGTCGGTCTAATGGCATTAATTTTACTAATGTGAATGGTAGTAATCCAGT 2749
Qy 2233 gaaacggtgtaataattacgaccaatagacagcacaggtacagttacgttggcattgac 2292
Db 2750 ACTGGCATTTGCAATATACCGCTCGCATTTA---CCAGAGATAAATTTGGCTTTGCTGGTCT 2806
Qy 2293 caaagtaattggtctcaccacgcctaaagctgacgctgggtagcgaacacaaatggttaatga 2352
Db 2807 GATGTGCGAGTTGATAC-----AACAAACCTTTATCTTGATCAAGACAAGCTA--CAA 2857
Qy 2353 ttggttattgagcaagtccttagcgtgacggttaacagacacacaaacacatcattaaagga 2412
Db 2858 GTTGGCAATTTAAGATTACCAACACTGGCATTAACGCGAGGTGGTAAAGCCATCAACGGG 2917
Qy 2413 ttgtcccaacacgctcgtacgttcaggtccaggtccaggtccgcaacacagcactggcgaat 2472
Db 2918 CTGTCCCAACACACTGCCTAGCATTCGCGATCAAAAGTAGCCGCAACATAGAACTGGGCAAT 2977
Qy 2473 acaatcgaagaaaaagacaaatcccaacgctgcagcattgtagtgctaaatcagggc 2532
Db 2978 ACAATCCAAGACAAAGACAAATCCAAACGCTGCCAGCATTAATGATATATTAAATACAGC 3037
Qy 2533 tttaacctaaaaataatggaagacaaagacttgcacacttgcacactgacaggttaatggc 2592
Db 3038 TTTAACTCTAAAAATTAATAACAACCCCATTTGACTTTGTCTCCACTTATGACATTTTGAC 3097
Qy 2593 ttatcgtgagcaatgcccacacgcccacagtaacttatgatgaagcacaacaaacaggt 2652
Db 3098 TTTGGCAATGGCAATGCCACCCACCCACAGTAACCCATGATACCCGTAAACAAAACCCAGT 3157
Qy 2653 aaagtggcgctgtagtgaattggtgtagaagaaacacttgaactgacagggcgaataatggc 2712
Db 3158 AAAGTGGTATATGATGTGAATGTGGATGATACAACCATTCATCTAACAGGCATGATGAC 3217
Qy 2713 aagaacaacttggtgcgaacacatcaaaactgacccgaagaaagacttaattggtaatgca 2772
Db 3218 AATAAAAACTTGGCGTCAAAACCCACCAAACTGAACAAAACAAAGTGTCTAATGGTAATACA 3277
Qy 2773 actacat-----ttagtaccgacgatgacacctgacctgtttaaagcaggtatctc 2823
Db 3278 GCAACTACTTTTAATGTTAACTCTTAGTGATGAAGATGCGCTTTGTAACGCCAAAGACATC 3337
Qy 2824 gccggcaactcaaacacccctagcgaggaatttcacacacacacaaagggcacagcaaaccc 2883
Db 3338 GCCGAAATCTAAACACCCCTAGCCAAAGGAATTCACACCACCAAAAGGCACACGACACACC 3397
Qy 2884 gccctacaaacctttaccgttaaaaaggtagatgataaagataagagctgtagcacccaaac 2943
Db 3398 GCCCTACAAACCTTTACCGTTAAAAAGGTAGATGAAAATAATATGCTGATGATGACGCCAAC 3457
Qy 2944 gccatcacggtgggttaagatggcaaggtggttaaggtcaaacaccttaaaactcaaaagt 3003
Db 3458 GCCATCACCGTGGGTCAAAAGAACCGCAATAATCAAGTCAACACCCCTAACACTCAAAAGT 3517
Qy 3004 aaaaacgctcttgattataaacacgacaaagatggtgacggttaccttggcattaacacc 3063
Db 3518 GAAACAGCTCTTAATATTAAACCCACCAAAATGTTACGGTTACCTTTGGCATTAACACC 3577
Qy 3064 caaagcggtcttaagacggcgagacaccactcttaaacacaaatgggttctctataaa 3123

Db	3578	ACAGCGGCTTAAAGCGGCAAAAGCACC---CTAAACGACGGTGGCTTCTATTAAA	4655	TTATCTGGGACATCCAAACTGCCAAAGGGGCAAGCCAAAGCAAACTCAGCAGGCTAT	
Qy	3124	aacacgcgtagtaaacgaacaaatccaagtcggtgctgctgagtgagtggttgccatggt-	Qy	4201	gtggatgctgtaggcaataaaggtcatctatgacagtacgcgataaacaagtactatcaagcc
Db	3635	AACCCACACTGGTAGCGAACAATCCAAAGTCGGTGTGATGGGTGAAGTTTGCCNAGGTT	Db	4715	GTGGATGCTCATGGCAATAAAGGTCACTATGACAGTACCGATAACAAGTACTATCAAGCC
Qy	3183	--Caataatgggtgtgtagtgctggtgctggtgacacaaactcgattaccagagatgaa	Qy	4261	aaaaatgtaggcacagttgataaaaccaaagaagttgccaagagacaacactggtcgcccaa
Db	3695	AATAAATAATGGTGTGTAGTGGCTGCTGATGGCACAACCTCGCATTACCAGAGATGAA	Db	4775	AAAATCATGCGCACAGTTGATAAAACCAAGAAAGTTGCCAAAGACAACTGTCGCCCAA
Qy	3241	atgggtttactgggactaatggctcaacttgataaaagcaaacccacactgaagcaaac	Qy	4321	gccaaaccccgatggcacatttggctcaaatgaatgtaaatcaaatcagtcatttaacaagaa
Db	3755	ATTGGCTTACTGGGACTAATGGCTCATTGATATAAAGCAAAACCCACCTAAGCAAAAGC	Db	4835	GCCCAAAACCCAGATGCGACATTTGGCTCAAAATGAATGTCAAAATCAGTCATTACAAGNA
Qy	3301	ggcatlaacgcaggtggttaaaagattacacacattcaatcaggtgagattgccaaaac	Qy	4381	caagtaaatgatgcataataaaagcaagggcatcaatgaagacacacgcttggtaagga
Db	3815	GGCATTAACGACGGTGTAAANAGATTACCAACATTCAATCAGGTGAGATTGCCCAAAAC	Db	4895	CAAGTAATGATGCCAATAAAAGCAAGGCATCAATGAAGACAAACGCTTTGTTAAGGA
Qy	3361	agccatgatgctgacagggcggaagatttatgatttaaaacgcgaactgaataaaa	Qy	4441	cttgaaaaagcgcgtctctgatacaaaaaacaaacacacacacacacacacacacacac
Db	3875	AGCCATGATGCTGTGACAGCGGCAAGATTATGATTTAAACCCGAACTTGAAACAAA	Db	4955	CTTGAAAAAGCGCTTCTGATAACAAACCAACCAACCCGCGAGTAACTGTGGGTGATTTA
Qy	3421	atcagcagtagtctccaaacagcacaaaactcattacacgaattctcagtagcagatgaa	Qy	4501	aatccgcttgcacaaacacacgctgacctttgcaggggatacagggcacaacggcttaaaaa
Db	3935	ATCAGCAGTACTGCCAANAACAGCACAAAACCTATTACAGAAATTCAGTAGCAGATGAA	Db	5015	AATGCCGTGGCCAAACACACGCTGACCTTTGCAGGGGATACAGGCACACGGCTTAAAAA
Qy	3481	caaggttaataactttacggttagtaaccccttactcaagttatgaacacctcaagacctct	Qy	4561	ctggcgagagactttgacccatcaaaaggtgggcaaacagacacacacacacacacacac
Db	3995	CAAGGTAAATAACTTACGGTTAGTAACCCCTTACTCCAGTTTATGACACCTCAAGACCTCT	Db	5075	CTGGCGGACACTTTTACCATCAAAAGTGGGCAAAACAGACACCAATTAAGTTAACCGTAAT
Qy	3541	gatctcatcactttgcagtgaaacggcattaccacacgaagtaataaaggtggtg	Qy	4621	aacatcgctgtgtagcaggttactgagcttctactgttcaaaccttgccaaagacctaacc
Db	4055	GATGTCATCACTTTGACGGTGAACAGCGCATTTACCAACCAAGGTAAATAAAGGTGGTG	Db	5135	AACATCGGTGTGTAGCAGGTACTGTATGGCTTCACTGTCAAACTTGCCAAAGACCTAAC
Qy	3601	cgtgtggcattgaccacaaacaaaggcttaaccacgcgttaagctgacgctgggtaaat	Qy	4681	aatcttaacagcgttaatgacggtggccacaaatgtagtaaaagggcgtctcttttga
Db	4115	CGTGTGGCATTTACCAAAACCAAGGCTTAACCAACGCTTAAGCTGACCGTGGGTAAAT	Db	5195	AATCTTAACAGCGTTAATGACAGGTGGCACCAAAATTTGATGACAAAGGGGTCTTTTGA
Qy	3661	aatggcaagcagtcattacacagcaaaatggtcaaaatggtacacagcaggaacga	Qy	4741	gacgcaaacggttcaagcacaagcaaacacccctgtgctcaagtggccaaatgggctggacctg
Db	4175	AATGCCAAGGCAATGTCATTGACAGCCCAAAATGCTCAAAATACCATCACAGGACTAAGC	Db	5255	GACTCAAGCGGTCAAGCCCAAGCAACACCCCTGTGCTTAAGTGCCCAATGGGTGACCTG
Qy	3721	aacactctagctaatgttaccaatgataaaggtagcgaacacacacacacacacacacac	Qy	4801	ggtggcaaggttcactcagtaattgtgggcaagggcacaagaataccgacgctgcgaatga
Db	4235	RACACTCTAGCTAATGTTACCAATGATAAAGGTAGCGTAGCGACACAGAACAGGGCAAT	Db	5315	GGTGGCAAGGTCATCAGTAATGTGGGCAAGGCACAAAGATACCGACGCTGCCAATGTA
Qy	3781	ataatcaaacgacgaagacaaaacccgtgcgcacagcattgtgtagtgtaagcgacgc	Qy	4861	caacagttaaacgaagtagcaactgttgggtct-----tggtaaatgataacgct
Db	4295	ATAATCAAGAGCAAGACAAACCCGTCGCCCGCCAGCATTTGTATGTGCTTAAGCGCAGGC	Db	5375	CAACAGTTAAACGAAGTAGCAACCTTGTGGGTCTTGGTAATGCTTGGTAATGATAACGCT
Qy	3841	tttaacttgaaggcaatggtgaagcgttgactttgtctccacttatgacacacgctcaac	Qy	4912	gacggcaatcaggttaaacatttccgcacatacaaaaaagaccccaaatccaggttcacatct
Db	4355	TTTAACCTGCAAGGCAATGGTGAAGCGGTGACTTTGTCTCCACTTATGACACCGTCAAC	Db	5435	GACGCAATCAGGTAAACATTTCCGACATCAAAAAGACCCCAATTCAGGTTTCATCATCT
Qy	3901	tttgcaaatggcaatatacacaacccgttaagtgacctatgatgacacaaagcaaacagct	Qy	4972	aacgcacactgatacaaacgagcaggttacttggcggtaaaaggttaataacgataaccgaa
Db	4415	TTTGGCGATGGCAATGCCACACCCGCTAAGGTGACCTATGATGACACAAAGCAAAACAGT	Db	5495	AACCGCATGTCTATCAAAAGCAGCAGGTACTTTGGCGGTAAAGGTAAATACCATACCGAA
Qy	3961	aaagtgtctatgatacgaatggtgatacaaacattgaagttaagataaaacactt	Qy	5032	aaacttgcacactggtggtgtacaaagtggcgctggaataaagcggcgaacgctaacggcgt
Db	4475	AAAGTGTCTATGATGTCATTTGGATGATACAAACCATTTGAAGTTAAAGATAAAAAGCTT	Db	5555	AACTTGGCCACTGGTGTATACAAAGTGGGCTGGATATAAAGACGGCAACGTAACGGCGAT
Qy	4021	ggcgtaaaac	Qy	5092	ttaagcaatgttgggttcaaaaccccaaaagatggcagcaaaaagccctgctcgccact
Db	4535	GGCGTAAAAACCAACCAACATTTGACAGTACTGGCAGAGGTGCTTAATAATTTGCCCTAAGC	Db	5615	TTAAGCAATGTTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGTCTGCCACT
Qy	4081	aatcaagcttactggcagtcgcttggtaagggcagtgatctggttgcctcatctaaacacc	Qy	5152	tataaccccgaggtcagacacacatagtgaccaaacaacccccgcagaagcattgacaga
Db	4595	AATCAAGCTACTGGCGATGGCTTGTCAAGGCCAGGTGATGCTGCTCATTAACACACC	Db	5674	TATAACCGCCGAGGTGACACCACTATTTCACCAACAAACCCCGCAGAAAGGCATTGACAGA
Qy	4141	ttatctgggacatccaaactgcacaaagggggcaagcgaagcgaacacactcagcaggtat	Qy	5212	ataaatgaacaaaggtatccgcttcttccatgtcaacgagtggaatcaagagcgttggtta
			Db	5735	ATAAATGAACAAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGACGCTGTGGTA

QY 3003 taaaaacggtcttgatattataaaacccgacaaagatggttacggttaccttggcattaaacac 3062
Db 3518 TGAACAGCGGTCTTAATATTAAACCGCAAAATATGGTACGGTTTACCTTTGGCATTAACAC 3577
QY 3063 ccaagcggctcttaaacgcccgcagacacacactctataaacaacatggcttgcctattaa 3122
Db 3578 CACAAGCGGTCTTAAGCGCGCAAAAGCACC--CTAAACGACCGGTGGCTTGTCTATTAA 3634
QY 3123 aaacccgctagtagtaacgaacaaatcccaagtcgggtgctgtagggcgtgaagttgcatgggt 3182
Db 3635 AAACCCCACTGGTAGCGAACAATCCAAGTCGGTGCCTGATGGCGCTGAAGTTTGCCAAGGT 3694
QY 3183 ---taataagtggtttgtagtgctgctgattgagtcagacaaactcgattaccagagatga 3239
Db 3695 TAATAAATAAGTGTGTAGGTGCTGGCAATTTGATGGCACAACCTCGCATTAACGAGATGA 3754
QY 3240 aattggcttactgggactaatggctcacttgataaagcaaacccacacctaagcaaga 3299
Db 3755 AATTGGCTTTACTTGGGACTAATGGCTCACTTGTATAAAGCAAAACCCCACTTAACCAAGA 3814
QY 3300 cggcattaaacgagtggttaaaagattaccaacattcaatcaggtgagattggccaataa 3359
Db 3815 CGGCATTAAACGAGTGGTAAAGAGATTACCAACATTCATCAGGTGAGATTGCCCAAA 3874
QY 3360 cagccatgactgctgacagggcggaagattatgatttaaaacccgaacttgaaaataa 3419
Db 3875 CAGCCATGATGCTGTGACAGCGGCAAGATTATGATTTAAATAACCGCAACTTGAACAA 3934
QY 3420 aatcagcagtagtctgcaaaacagacaaaactcaatcacgaattctcagtagcagatga 3479
Db 3935 AATCAGCAGTACTGCCAAACAGCACAAACTCATTTACAGCAATTCAGTAGCAGATGA 3994
QY 3480 acaaggtaataacttaccggttagtaaccttactccagttatgacacctcaaaagacctc 3539
Db 3995 ACAAGGTAATAAATTTACGGTTAGTAACCCCTTACTCCAGTTATGACACCTCAAAAGACCTC 4054
QY 3540 tgatgtacacacttgcaggtgaaacgcttaccacacgaagtaataaagtggtggt 3599
Db 4055 TGATGTATCATCACCCTTGCAGGTGAACCGGCAATTTACCACCAAGGTAATAAAGGTGTGGT 4114
QY 3600 gcgtgtggcctattgacaaaacaaaggcttaaccacgcctaaagctgacgcgtgggtaataa 3659
Db 4115 GCGTGTGGCATTCACCAACCAAGGCTTAACACAGCCCTAAAGCTGACCGTGGGTATTA 4174
QY 3660 taatggcaaggcattgtcattaaacagccaaaatgggtcaaaataaccatcacagactaag 3719
Db 4175 TAATGGCAAAAGGCATTTGTCATTGACGCCAAATATGGTCAAAATACCATCACAGGACTAAG 4234
QY 3720 caacacttagctaatgttaccatgataaaggtagcgtaacgacacacagaaagggcaa 3779
Db 4235 CAACACTTAGCTAATGTTACCAATGATAAAGGTAGCGTAGCGTACGCCACACAGAACAGGCGAA 4294
QY 3780 tataatcaaacgacgaacaaacccgtgcgcgcagcattgttgcgtgtaacgcagcag 3839
Db 4295 TATTAATCAAGACCAAGCAAAACCCGTGCCGCCAGCATTTGTTGATGTGCTTAAGCGCAGG 4354
QY 3840 ctttaacttgcgaaggcaatggtagcgggttgacttgcctccacttatgacacccgtcaa 3899
Db 4355 CTTTAACCTTGAAGGCAATGTGAAGCGGTTGACTTTGTCTCCACTTATGACACCGCTCAA 4414
QY 3900 ctttggcaatggcaataaccacccgcttaagtgacctatgatacacaagcaaaacag 3959
Db 4415 CTTTGGCGATGGCAATGGCCACCCCGCTTAAGGTGACCTTATGATGACAAACCAAAACCCAG 4474
QY 3960 taaagtggtctctatgctcaattgtggaatgatacaacacttgaagttaagataaaaaact 4019
Db 4475 TAAAGTGGTCTATGATGTCATATGTGGATGATACAAACCATTTCAAGTTAAAGATAAAAACT 4534
QY 4020 tggcgttaaaacaccacacattgaccagtagctggtgacaggtgcttaataatttgccttaag 4079
Db 4535 TGGCGTAAAAACCAACCATTTGACAGTACTTGGCACAGGTGCTAATAAATTTGCCCTTAAG 4594
QY 4080 caatcaagctactggtgcgtgtgtcaaggccagtgatctgttgcctatcaaacac 4139

Db 4595 CAATCAAGCTACTGGCGATGCGTGTCAAGGCCAGTGATATCGTTGCTCATATAACAC 4654
QY 4140 cttatctggcagactcacaactgcaaaaggggcgaagccaaagcgaacaactcagcaggcta 4199
Db 4655 CTTATCTTGGCGACATCCAACATGCGCAAAAGGGCAAGCCAAAGCAACTCAGCAGGCTA 4714
QY 4200 tgtgtagtgcgtggaataaagggtcactctatgacagttaccgatacaaaagtactatcaagc 4259
Db 4715 TGTGGATGCTGATGGCAATAAAGGTCACTATGACAGTACCGGATAACAAGTACTATCAAGC 4774
QY 4260 caaaatgtagtcagacttgataaaaacaaagaagttgcgaagacaaactggtcgccca 4319
Db 4775 CAAAATGATGGCACAGTTGATAAAACCAAGAGTTGCCAAAGACAAACTGGTCGCCCA 4834
QY 4320 agcccaaacccagatggcacattggctcaaatgaatgcataatcagtcattaaacaaga 4379
Db 4835 AGCCCAAAACCCAGATGGCACATTTGGCTCAATGAATGCAATCAGTCATTAAACAAGA 4894
QY 4380 acaagtaaatgatgccaaataaaaagcaaggcatcaatgaagaacacgaccttggtaaagg 4439
Db 4895 ACAAGTAAATGATGCCAATAAAGCAAGGCATCAATGAAGACAACGCGCTTTGTTAAAGG 4954
QY 4440 acttgaaaaacgcttctctgataaaaacaaacccaaacgcccagtaactgtgggtgattt 4499
Db 4955 ACTTGAAAAAGCCGCTTCTGTATAACAAACCAAAACGCCGACGTAACGTGGGTGATTT 5014
QY 4500 aatgcgcgttgcacaaacacccgcgtgaccttgcagggggatacagacacacgctcaaaaa 4559
Db 5015 AAATGCCGTTGCCCAACACCGCTGACCTTTTCAGGGGATACAGCACCAACGGCTTAANA 5074
QY 4560 actggggagacttggaccataaaaggtyggcaaacagacaccaataaagctaaacgataa 4619
Db 5075 ACTGGCGAGACTTTGACCATCAAAAGGTGGCAACACAGACACCAATAAGCTAAACGGATA 5134
QY 4620 taacatcggtgtgtagcagtagctgactgctcactcaactcaactgcaagacactaac 4679
Db 5135 TAACATCGGTGTGGTAGCAGGTACTGATGGCTTCACTGTCAAACTTGCCAAGACCTAAC 5194
QY 4680 caacttaacagcgttaatgcaggttggcaccaaaattgatgaaaagcactctctttgt 4739
Db 5195 CAATCTTAACAGCGTTAATGTCAGGTGGCCACCAAAATTCATGACAAAGCGGTGCTCTTTGT 5254
QY 4740 agacgcaaacggttcaagccaaagcaaacacccctgtgttaagtgcataatgggtggaact 4799
Db 5255 AGACTCAAGCGTCAAGCAAAAGCAAAACACCCCTGTGCTAAAGTGCCAATGGGCTGAGCT 5314
QY 4800 gggtggcaagtcactcagtaagtgggcaaaagcacaagaataccgacgctgccaatgt 4859
Db 5315 GGGTGGCAAGGTCAATGATGTGGGCAAAAGGCAAAAGATACCGACGCTGCCAATGT 5374
QY 4860 acaacagttaaacgaagtacgcaacttgttgggtct-----tggtataatgaacgc 4910
Db 5375 ACAACAGTTAAACGAAGTACGCAACTTGTGGTCTTGGTAACTGCTGATGATAACGC 5434
QY 4911 tgacggaactcaggttaaacatgacgacatacaaaaagacccaaattcaggttcatc 4970
Db 5435 TGACGGCAATCAGGTAAACATTTGCCGACATCAAAAAGACCCAAATTCAGGTTTCATC 5494
QY 4971 taacgcaactgctcaaacgagcagcagcttacttggcgtaagtaataacgataccga 5030
Db 5495 TAACCGCACTGTCAATCAAGCAGGCGGCTACTTGGCGGTAAAGGTAAATACGATACCGA 5554
QY 5031 aaaaacttccactggtgtgataaagtggcggtgataaaaagcgaacgctaaacgga 5090
Db 5555 AAAACTTCCACTGCTGCTGATACAAAGTGGCGTGGATAAAGACGCAACGCTAACGGCA 5614
QY 5091 ttaagcaatgttgggtcacaaccccaaaaagatggcgcaaaaaagccctgctgcac 5150
Db 5615 TTTAAGCAATGTTTGGGTCAAAACCCAAAGATGGCAGCAAAAAGCCCTGCTGCCAC 5674
QY 5151 ttataacgcccaggttcagacacactatgtgacacaaccccgagagcattgaag 5210

Db	1256	GGCAGTGGGTCTTCGCGCCACACGCCGAGGGCCAAATCTACAAATGCTATTTGGTTCTGTGATGC	1311
Qy	762	aaaagctaaacgcttggccagcagacagccattgggtggaataactctagtttaatttgggtcg	821
Db	1316	AACATCTAGCTCGTTGGGAGCGCATAGCCCTTGGTGACGGTACTCGTCTCAGCTACAGGG	1375
Qy	822	aggogttgccctaggtttggttctcagatactcttgatagggataataataacagatgccag	881
Db	1376	CAGTATTGCCCTAGGTCAAGGTTCTGTGTCTACTCAGAGTGATATAATATTC---TAGACC	1432
Qy	882	tgcctatgtaccactaggtaaaacgtttagcagacagcagataaaaagccaccgcgcagggtga	941
Db	1433	GGCCTATACACAAATPACCCAGGCATAGACCCCAAGTTTCAAGCCAC---CAATAATAC	1489
Qy	942	tctcaggatatatttccatctggaatagtaataataataatagcagatcacaggcgtaa	1001
Db	1490	GAAGCGGGTCCATTTCCTATTGGTAGTAA-----CTCTATCAACCGTAA	1534
Qy	1002	aatcatcaatgtcgtgcgggttcttcgggatacogatcggtcctaattgtgcacagcttaa	1061
Db	1535	AATCATCAATGTCGGTGCAGGTGTTAATAAAMACCAGTCGCTCAATGTGCGACAGCTAGA	1594
Qy	1062	attggtggaggaactggcta---atcgtaaaattacttttaagggtgatggtgaacaataa	1118
Db	1595	AGCGGTGGTGAAGCTGGGCTTAAGAGCGTAGAATTACTTTTCAGGGTGATG-----ATAA	1648
Qy	1119	tagcaatagcgtagaagaggtttgggcaactactttaactattaaagtgatgcacagac	1178
Db	1649	CAGTACTAGCTAAAATAGGTTTGGATAATACTTTTAACTATTAAAGGTGGTGCAGAGAC	1708
Qy	1179	caacgcattaaccgaagctaaacatcggtgtggtgaacagatggcaa-----tggtctgaa	1232
Db	1709	CAACGCATTAACCGATAATAATATCGGTGTGTGTAAAGAGGCTGATAATAGTGGTCTGAA	1768
Qy	1233	agttaaactgtctaaagactgactggattgaccagtgtctcc-----gc	1277
Db	1769	AGTTAAACTTTGCTTAAACCTTTTAAACAATCTTACTGAGGTGAATACAACATACATTAATATGC	1828
Qy	1278	taccacaataatcacgcttagtaataaccaacaacaacgccgagctacaaagcgtg	1337
Db	1829	CACAAACCACATTTAAGGTAGGTAGTAGTAGTAGTACTACACTGAAATTTAGTGATAG	1888
Qy	1338	ttgaccttttagccca-----atacaggtgtacaaaaacagataaaacgctctacagcat	1391
Db	1889	TTTAAACCTTTACCCAGCCCAATACAGGCAGTCAAAAGCACAAAGCAAAACCGTCTATGGCGT	1948
Qy	1392	tgatgattgaadttaactaaatgatatagtaataagtaatacgaactaaagttactactcgtat	1451
Db	1949	TAAATGGGGTGAAGCTTTTACTAAATTAATGCAGAAACACACAGCACTACGGCAC/TACTCGTAT	2008
Qy	1452	taccaaaaagaaaattggttttgcgtggtactaatgatggagtgtatgaagaagcaaacctta	1511
Db	2009	TACCAGAGATAAAATTTGGCTTTGCTCGAGTGGTGATGTTGATGAAAAMAAACAGCACCATTA	2068
Qy	1512	tcttgacaacgaagaagctaaaagttggcaacagcaccctaacacagttggtagcttgactgt	1571
Db	2069	TTTGGATAAAAAACAATTAAGTGGTAGTGTGTCAATTTACCATAGACAATGSCATTGA	2128
Qy	1572	taataacacacactggttaaataacaaatccaagtcggtgctaataggcattaaatgtgccac	1631
Db	2129	TGCAGGTAATAAAAAGATCATGTAATCTTGCCAAAAGGTAGCAGTGTCTAACGATGGCGTTAC	2188
Qy	1632	agtcgctaaataatgtgcgaataacctcagcaacagtcggcactgctcgtattaccgaaga	1691
Db	2189	CATCGAACACCTCAAAAGCGCCNAGCCTACTTTTAAACGCGAGGCGTGGCATCTAGTGTAC	2248
Qy	1692	gaaaattggttttgcgttactaatgatggagtgtatgaacaagcaccataatttggataa	1751
Db	2249	ACCTACTGAAATATCAGTTGATGCTAAGAGTGGCAATGTTACCGCCCCAACCTTACAACA-	2307
Qy	1752	agaacgacttaagtggtggtctgttgtaaataccacagatagttggtattaatgctgtdaa	1811
Db	2308	-----TTGGGCTGAAAACACCACCGAGCTTAACAGTGTGGCTACTGTGATAAATTTAG	2359

Qy	1812	tcaaaagattaccggacttactaaatggtatagcaaataccgatccggtttaccatcaaa	1871
Db	2360	TGTTAAGGGTAGTGGTACGAACAATAGCTTAGTTACCGCCGAACATTTGGCAAGCTATCT	2419
Qy	1872	gctcaaaagacgcgaagcctactttaaaccgagcgagatggcatcagatttaataagtaata	1931
Db	2420	AAATGAAGTCAATCGAACCGGCTGCACAGTGCCTCTCAAAAGCTTTACCGTTAAAGAAGAAGA	2479
Qy	1932	cggggactctagttgtagtagtggcaatataccaccaccaactataaacattagcgtgaa	1991
Db	2480	CGATGATGACG-----	2509
Qy	1992	aaccactaaagcttaaacagtaaatggcaccagtggttaataataaatttagttagtaatgc	2051
Db	2510	TAAAGATATCGACAAAAAATGCGGGCGACGTACGATCTTTAAAACTCAAAAGGTAAAAACGG	2569
Qy	2052	tcatgataacaatagcttggttacgcgcgaagattggcagactatctaaataaagtcga	2111
Db	2570	TCTAACGGTTGCTTACCAAAAAAGATGGTAGCTTACCTTTGGGCTTAGCCAAGATAGCGG	2629
Qy	2112	tgaacgcgtgcacagtgctctaccaagctttaaagtcacaaacgcgtgataaagaacaa	2171
Db	2630	TCTGACCATTTGGCAAAAGCACCCCTAAACAAACGATGGCTTGACTGTTAAAGATACCAACGA	2689
Qy	2172	cgccatcacgcgtgggttaaagatacacaacgcgcgaagaccttcaacaccttaaaactcaaaag	2231
Db	2690	ACAAATCCAAAGTCGGTGCTTAATGGCATTTAAATTTACTAATGTGAATGTTAGTATCCAGG	2749
Qy	2232	tgaacgcgtggttatattacgaccaatagaccacaggtacagttaccttggttcattga	2291
Db	2750	TACTGGCATTTGCAAAATACCGCTCGCATTA--CCAGAGATAAAATTTGGCTTTGGTGGTTC	2806
Qy	2292	ccaaagttaatggtctcaacacgcctaagctgacctgggtagcgaatacaaatgtaatcg	2351
Db	2807	TGATGGTGCAATTTGATAC-----AAACAACCTTTATCTTGATCAAGACAAGCTA--CA	2857
Qy	2352	attggttattgagcaagtcctagcgtgcagcgttaacgacacaaaacatctaatgaag	2411
Db	2858	AGTTGGCAATGTTAAGATTACCAACACTGGCATTAACGCAGGTGGTAAGCCATCACAGG	2917
Qy	2412	attgtccccaacactgcctagcattgccaggtccaagtgccgcaacatgacatcggtgcaa	2471
Db	2918	GCTGTCCCAACACTGCTAGCATTTGCCGATCAAAAGTAGCCGCAACATFAGAAGCTGGGCAA	2977
Qy	2472	tacaatcgagaaagaaagacaaaatccaacgctgcagcgttaacgacacaaaacatctaaag	2531
Db	2978	TACAATCCAAGACAAGACAATFCCAACGCTGCCAGCATTAATGATATATTTAAATACAGG	3037
Qy	2532	ctttaacctaataataatggcaagacaagaactgtctccactatgacactgttga	2591
Db	3038	CTTTTAACCTAAAAAATAATCAACCCCAATGACTTTGTCTCCACTTATGACATTTGTGA	3097
Qy	2592	ctttatcgatggcaatggccaccaccgccacagtaacttatgatgaagccaatacaaacag	2651
Db	3098	CTTTTGCCAAATGGCAATGCCACCACCGCCACAGTAACCCATGATACCGCTTAACAAAACAG	3157
Qy	2652	taaagtggcgtatgatgtgaatgtgagaaaaaccattgaactgcagcagcgataatgg	2711
Db	3158	TAAAGTGGTATATGATGTGAATGTGATGATACAACCAATTCATCTTAACGACGACTGATGA	3217
Qy	2712	caagaacaacttggcgtcaaaaccatcaactgaccgaacaagaactactaatgtaatgc	2771
Db	3218	CAATTAANAACCTTGGCGTCAAAACCCACCAAACTGAACAAACAAAGTGGCTAATGTGTAATAC	3277
Qy	2772	aactacat-----	2822
Db	3278	AGCAACTTAACCTTTAATGTTAACTTAGTCATGATGAAGATGCCCTTGTTAACGCCAAAGACAT	3337
Qy	2823	cgccggcaatctaaacacccctagccgaggaaattcacaccaccaagggcagcaaacac	2882
Db	3338	CGCCGAAATATCTAAACACCCCTAGCCAAGAAATTCACACCAACCAAGGCACGACGACAC	3397

QY	2893	cgccctcaaacctttaccggttaaaaaggtagatgaaatgataagctgtagcaccaaa	2942
DB	3398	cgccctcaaaacctttaccggttaaaaaggtagatgaaatgataagctgtagcaccaaa	3457
QY	2943	cgccatcaccttggttaaaagtgacacaagtggttaaagtcacacaccttaaaactcaaa	3002
DB	3458	cgccatcaccttggttggttcnaaaagacgcaaatatcaagtcaacaccttaacactcaaa	3517
QY	3003	taaaacccggtcttgatataaaacccgacaagaatggttacggttaccttggcattcaac	3062
DB	3518	tgaanaacccggtcttgatataaaacccgacaagaatggttacggttaccttggcattcaac	3577
QY	3063	ccaaagcgttttaaacgcccgcagcaccacctcttaacacaaatggtgtgtctattaa	3122
DB	3578	cacaagcgttttaaacgcccgcanaaagcacc--ctnaacgacggtggcttgtctattaa	3634
QY	3123	aaacacccgctagtaacgaacaaaaatccaaagtcggtgctgtagtggttgaaagtttgcattg	3182
DB	3635	aaacacccgctagtaacgaacaaaaatccaaagtcggtgctgtagtggttgaaagtttgcattg	3694
QY	3183	---taataatgctgttgtagtgctggtgcatgtaggcacaactcgcattaccagagatga	3239
DB	3695	taataataatgctgttgtagtgctggtgcatgtaggcacaactcgcattaccagagatga	3754
QY	3240	aattggcttactgggactaattggtcactctgataaaagcaaaacccacctaagaaga	3299
DB	3755	aattggcttactgggactaattggtcactctgataaaagcaaaacccacctaagaaga	3814
QY	3300	cggcattaacgaggttgtaaaagattaccacactcaatcaggtgagattggcaca	3359
DB	3815	cggcattaacgaggttggttaaaacaaataccacaactcaatcaggtgagattggcaca	3874
QY	3360	cagccatgatgctgtgacagggcggaagatttatgatttaaaacccgaaacttgaaataa	3419
DB	3875	cagccatgatgctgtgacagggcggaagatttatgatttaaaacccgaaacttgaaataa	3934
QY	3420	aatcagcagtagtgcacaaacagcacaaaactcattacgaatcttcagtagcagatga	3479
DB	3935	aatcagcagtagtgcacaaacagcacaaaactcattacgaatcttcagtagcagatga	3994
QY	3480	acaaggttaataactttacggttagtaaccccttactccagttatgacacctcaaaagac	3539
DB	3995	acaaggttaataactttacggttagtaaccccttactccagttatgacacctcaaaagac	4054
QY	3540	tgatgcatcaccttgcaggtgaaacggcattaccaccaaggtaaaataaagggttggt	3599
DB	4055	tgatgcatcaccttgcaggtgaaacggcattaccaccaaggtaaaataaagggttggt	4114
QY	3600	gcgtgtgggcattgacaaaccaaaggcttaaacacgcctaaagctgacggttggttaataa	3659
DB	4115	gcgtgtgggcattgacaaaccaaaggcttaaacacgcctaaagctgacggttggttaataa	4174
QY	3660	taatggcaaggcattgattacacgcaaaaatggtcaaaaataccatcacaggactaa	3719
DB	4175	taatggcaaggcattgattacacgcaaaaatggtcaaaaataccatcacaggactaa	4234
QY	3720	caacactagctaagtgttaccatgataaaggtagctgcacacgaacacagggca	3779
DB	4235	caacactagctaagtgttaccatgataaaggtagctgcacacgaacacagggca	4294
QY	3780	tataatcaaaagcagaacaaaacccgtgcgcagcattgttgatgctgaagcgcagg	3839
DB	4295	tataatcaaaagcagaacaaaacccgtgcgcagcattgttgatgctgaagcgcagg	4354
QY	3840	ctttaacttgaagcaatggtgaagcgggtgactttgcttccacttatgacaccgtcaa	3899
DB	4355	ctttaacttgaagcaatggtgaagcgggtgactttgcttccacttatgacaccgtcaa	4414
QY	3900	ctttgccaatggcaataccacacccgtcaaggtgacctatgtagcaacagcaaaac	3959
DB	4415	ctttgccaatggcaatgccacccgtcaaggtgacctatgtagcaacagcaaaac	4474
QY	3960	taaaagtggtctatgtagtcaatgtggaatgatacaacattgaagtttaaaagataaaa	4019

Db	4475	 TAAAGTGGCTATATGATGTCAATGTGGATGATACACCATTTGAAGTTTAAAGATAAAAAACT	4534
Qy	4020	 Tggcgtaaaaccaccacattgaccagtgactggcacaggtgctaaataaatttgccttaag	4079
Db	4535	 TGGCGTAAAAACCACCATTTGACCACTACTGGCACAGGTGCTTAATAAAATTTGCCCTAAG	4594
Qy	4080	 caatcaagctactggcgatgcgtgttcaaggccaagtgtatcgttgcctcatctaacaac	4139
Db	4595	 CAATCAAGCTACTGGCGATGGCGTGTGTCAGGCCAGTGATATCGTTGCTCNTCTTAACAC	4654
Qy	4140	 cttatctggacatccaaactgcctaaaggcgcaagccaagcgcaacaactcagcaggtta	4199
Db	4655	 CTTATCTGGCGACATCCAAACTGCCAAAGGGCGCAAGCAAGCGAACAACCTCAGCAGGCTA	4714
Qy	4200	 tgtgatgctgtagtgcataaaggttcattctgacagtaccgataaacaagtactatcaagc	4259
Db	4715	 TGTGGATGCTGATGTCATTAAGGTCATCTATGACAGTACCGATTAACAAAGTACTATCAAGC	4774
Qy	4260	 caaaaatgatgacacagttgataaaacaaagaagttgccaagaacaaactggctgcgcca	4319
Db	4775	 CAAAAATGATGCGCAGAGTTGATTAACAAACAAAGAGTTGCCAAGACAACTGGTGCGCCA	4834
Qy	4320	 agcccaaacccagatggtgcacattggctcgaattgaattcgaatcgaatcattacaaga	4379
Db	4835	 AGCCCAAAACCCAGATGGCACATTTGGCTCAAAATGAATTCAAATCAGTCAATTAACAAGA	4894
Qy	4380	 acaagttaatgatgccaataaaaaagcaaggcatcaataagaacacagccttgttgaagg	4439
Db	4895	 ACAAGTAAATGATGCCCATTAATAAGCAAGGCATCAATGAAGACAAACGGCTTTGTTAAGG	4954
Qy	4440	 acttgaagaagccgtctctgtatacaaaacccaaacccgcgcagtaactgtgggtgattt	4499
Db	4955	 ACTTGAAAAAGCCGCTCTGTATACAAAAACCAAAAACGCCGCGAGTAACTGTGGGTGATTT	5014
Qy	4500	 aaatgccgttggcccaaacaccgcctgacctttgcaggggatacagcagcaacagcttaaaa	4559
Db	5015	 AAATGCCGTTGGCCAAACACCCGTGACCTTTTCAGGGGATACAGGCACAAACGGCTTAANA	5074
Qy	4560	 actggcgagactttgacctaaagggtgggcaaacagacaccaataagctaacgcgtataa	4619
Db	5075	 ACTGGGCGAGACTTTGACCATCAAAAGTGGGCAACACAGACACCAATTAAGCTAACCGATAA	5134
Qy	4620	 taacatcggttgtgtagcaggtactgatggcttcaactgtcaaaacttgcgaagacctaac	4679
Db	5135	 TAACATCGGTGTGTTAGCAGGTACTGATGCTTTCATGTCAAACTTGCCTAAAGACCTTAC	5194
Qy	4680	 caacttaacaggttaatgcaggtggccacaaaattgatgaaaaggcatctctttgt	4739
Db	5195	 CAATTTTAAACAGCGTTAATGCAGGTGGCACCAAAATTGATGACAAAAGGCGTGCTTTTGT	5254
Qy	4740	 agacgcaaacgtcaagccaaagcaaacacccctgtgctaagtgcacaaatgggctggacct	4799
Db	5255	 AGACTCAAGCGGTCAAGCCAAAGCAAAACACCCCTGTGCTTAAGTGCCAAATGGGCTGACCT	5314
Qy	4800	 gggtggcaaggtcatcagtaattgtgggcaagcacaagaatcagcagcgtgcctaagt	4859
Db	5315	 GGGTGGCAAGGTCTATCAGTAATGTGGGCAAGGCAAAAGATACCGACGCTGCCAATGT	5374
Qy	4860	 acaacagttaaacgaagtacgcaactgtttgggtct-----tggtaagtataaacyc	4910
Db	5375	 ACAAACAGTTAAACGNAGTACCCAACTTTTGGGTCTTGGTAAATGCTCGTAAATGATAACGC	5434
Qy	4911	 tgacggcaatcaggtaaacattgcgcacatacaaaaaagaccaaaatcaggttcatc	4970
Db	5435	 TCAGCGCAATCAGGTAAACATTTGCCGACATCAAAAAAGACCCAAATTCAGGTTTCATCATC	5494
Qy	4971	 taacgcactgtcatcaagcagcagcaggttacttggcggttaaaagdttaataacgcataccga	5030
Db	5495	 TAACCGCATGTGTATCAAAGCAGGCACGGTACTTTGGCGTTAAAGGTAAAGCTAATTAACGATACCGA	5554
Qy	5031	 aaaacttgccaactgggtgttacaagtgggcgttgggataaagacggcaacgctaaccgca	5090


```
Db 5555 AAAACTTGCCACTGGTGGTATACAAAGTGGGGTGGATTAAGACGGCAACGCTAACGGCGA 5614
Qy 5091 ttaagcaattgttgggtcaaaaccccaaaagatggcagcaaaaaagccctgtctcgccac 5150
Db 5615 TTTAAGCAATGTTGGGTCAAAACCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCCAC 5674
Qy 5151 ttataacgcgcgcaggtcagaccactatgtgacccaacaccccgacagaagccattgacag 5210
Db 5675 TTATAACGCCGCAGGTGAGACCAACTATTGACCAACAACCCCGCAGAACCCATTGACAG 5734
Qy 5211 aataaatgaacaaggtatccgctcttccatgtcaacgatggcaatcaagagcctgtggt 5270
Db 5735 AATAAATGACACAGGTATCGCTTCTTCCATGTCAACGATGGCANTCAAGAGCCTGTGGT 5794
Qy 5271 acaagggcgtaagggcattgactcaagctgcctcaggcaagcactcagtgagtagttt 5330
Db 5795 ACAAGGGCGTAACGGCATTTACATCAAGTGCCTCAGGCAAGCACATTCAGTGGCGATAGTTT 5854
Qy 5331 ccaggccaagggcagatggtgaagccgcggttgccaataggcagacaaaacccaagcaggcaa 5390
Db 5855 CCAGGCCAAGGCAGATGGTGAAGCCGCGTGGCCATAGGCAGACAAAACCCCAAGCAGGCAA 5914
Qy 5391 ccaatccatcgcatcggtgataacgcacaagccacggcgcatcaatccatcgccatcgg 5450
Db 5915 CCATTCATCGCCATCGGTGATACGCACAAAGCCACGGGGGATCAATCCATCGCCATCGG 5974
Qy 5451 tacaggcaattggttagcaggtgaagcactctggtgcatcgggacccagcagcactgtaa 5510
Db 5975 TACAGGCAATGTGTAGCAGTAAAGCACTCTGTGGTGCATCGGCACCAAGCAAGCACTGTTAA 6034
Qy 5511 ggcgtataacagttacagttgggttaatacaaaccaagtttaccgatccactcaaacaga 5570
Db 6035 GGCTGATAACAGTGTGGGTAAATAACAACCACTTTACCGATGCCACTCAAAACCGA 6094
Qy 5571 tgccttgggtgggcaataaacatcacctgacccgaagtaactcgtgtgcttaggttc 5630
Db 6095 TGCTTTGGTGGGCAATTAACATCACCGTGACCGAAGTAACCTCGGTGCTTAGGTTT 6154
Qy 5631 aaactctgccatcagtcgagggcacacacgcagggcacacaagccaaaatactgacggcac 5690
Db 6155 AAACCTTGCCATCAGTCAGGCACACACGCGAGGCACACAAGCCAAAATACTGACGGCAC 6214
Qy 5691 agcaggttacaacacacacagcaggtgccaacaggttacaggttaagagccttgcggacaaac 5750
Db 6215 AGCAGGTACAAACCAACCAAGCAGGTGCAACCGGTACCGTTAAAGGCTTTGCTGGACAAAC 6274
Qy 5751 ggcggttgggtgcggtccggtgctcaggtgctgacggcagccggtatccaaaatgtggc 5810
Db 6275 GGCGGTGGTGGGCTTCCGTGGGTGCTCAGGTGCTGACGGCGGTATCCAAAATGTGGC 6334
Qy 5811 agcaggtgagtgccagtcaccagcaccgatgcggtgcaatggtagccagttgtacaagc 5870
Db 6335 AGCAGGTGAGTGCAGTCACACACCGATGCGGTCAATGGTAGCCAGTTGTTACAAAGC 6394
Qy 5871 caccacaagcatgccaacgaacaaatagcttgacatcgatccacacaaaacagaaa 5930
Db 6395 CACCAAAAGCATTCACCAACGCAACCAATGAGCTTGACCATGCTATCCACCAAAACAGAAA 6454
Qy 5931 taaagcaatgacaggtatttcacagcgatggcgatggcgccatgccaacagcctacat 5990
Db 6455 TAAGGCCAATGACAGGATTTTCATCAGCGATGGCGATGGCGTCCATGATCCACAAAGCCTACAT 6514
Qy 5991 tccctggcagatccatggttaccgggggtattggccaccacaaacggtagcgggtggc 6050
Db 6515 TCCCTGGCAGATCCATGTTTACCGGGGGTATTGGCCACCAACGATCAAGGTGCGGTGGC 6574
Qy 6051 agtgggactgcgaagctgcgaataatggtcgaatgggtatttaaaatcaaatggttcagc 6110
Db 6575 AGTGGGACTGCTCGAAGCTGTGGATAATGGTCAATGGGTATTTAAAAATCAATGTTTCAGC 6634
Qy 6111 cgatacccaagggccatgtagggcgcgagttggtggtgaggttttaccattt 6159
Db 6635 CGATACCCAAAGGCCATGTAGGGGGCGCAGTTGGTGGCAGGTTTTCACATT 6683
```

RESULT 6

```
AX067457 LOCUS AX067457 62909 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 32 from Patent WO0078968.
ACCESSION AX067457
VERSION AX067457.1 GI:12545077
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 62909)
AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 32 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
source
1..62909
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 18449 a 13798 c 12744 g 17918 t
ORIGIN
Query Match 64.7%; Score 3982; DB 10; Length 62909;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 5068; Conservative 0; Mismatches 785; Indels 355; Gaps 26;
Qy 1 atgaatcacatctataagtcattctttaacaaagccacagggcacatttatggcgtggcg 60
Db 42630 ATGAATCACATCTATAAGTCATCTTTTAAACAAAGCCACAGGCACATTTATGGCCGTGGCG 42689
Qy 61 gaatatccaaatccacagtcacggggggg-tagctgtgtctacagggcaagttggcag 119
Db 42690 GAATATGCCAANTCCACAGCAGCGGGGGGATAGCTGTGCTACAGGGCAAGTTGGCAG 42749
Qy 120 tgcacgacttaagcttggccgctattggcgcgctcgtctcctcgtgacgtgctgac 179
Db 42750 TGTAATGCACCTGAGCTTTGGCCGCTATTGGCCGCTCGCTGCTCGTGCATCGGTGGCAG 42809
Qy 180 gctcaatggcagtccttatgctcaacaaatctaccagaatcgaaatctggtcaaacaaa 239
Db 42810 GCTCAGTGGCAGTGCCTTATGCTCAA-----GGTATCGCAATTGGTCAA----- 42852
Qy 240 caagataaacacacgctgaagcgatgccctagcacaggtgaagcattccattgcttt 299
Db 42853 ----CAAAACAAACACTCAAAAGGCGATGCCGAGCGACAGGTAATCAATCCGTTGCTCT 42908
Qy 300 tggtagctcttcttaagggcacagggctctcaagctattgctatcggtagtgcaaacccaga 359
Db 42909 TGGTGAAAAGACCTGTTCGAAGTGGCAGTCAAGCCATCCCATCGGTGTAGCCAAACAA 42968
Qy 360 tcctaataatgggtgaatggtlaatgtaggttcccaacgccaagaagtaacagagtcacatcgc 419
Db 42969 ATCAAAATAATAGTA-----AAGTATCAGGTGCCAACGCCAACCCAAAGGTGAAGAGTCCATCGC 43022
Qy 420 catcgttggtgatgtattggtgaggtgagtcctcattccattccatcggtgagtcatt 479
Db 43023 CATCGGTGATGTATTGGCTCAGGGTCTCGCTTCGATTGCCATCGGTGATGATGACTT 43082
Qy 480 atatttgctcaagaa-----tcttgatctgaagaataatttcacaaacttattcatgg 533
Db 43083 ATATTGGAGAGTAAAGGCTTGTATACACTCCACAGAGAGATTAAAAAATTTATGAAA 43142
Qy 534 ccataaataataaaaaaacatac---aaacctcaacgatggtaaaatcaaatatcgacg 590
Db 43143 ACATTTCAGAAATTAACAGACAGATACGAAAACCTGGATCCCAATAGTAATCAAAAATATTATACA 43202
Qy 591 cacaagacacaagggcacgccactactcagttgggaacccatgctcatgcacaggggca 650
Db 43203 CAAAAAGCACAAGGACACGCCAGTATTTCAGTGGGAACCACTGGCACAGGCGAGGGTCA 43262
```


QY	6072	gataatgggtcaatgggtattataaatcaatggttcagccgataccaaggccatgtagg	6131
Db	48444	GGATATGGTCAATGGTATTTAAATCAATGGTTCAGCGATACCAAGGCATGTAGG	48503
QY	6132	ggcggcagttggtgcagggttttcacitt	6159
Db	48504	GGCAGCAGTTGGTGCAGGTTTTCACTTT	48531
RESULT 7			
AX079914	AX079914	5976 bp	DNA
LOCUS	Sequence 2 from Patent WO0107619.		PAT 22-FEB-2001
DEFINITION	AX079914		
ACCESSION	AX079914.1	GI:13159437	
VERSION			
KEYWORDS	Moraxella catarrhalis.		
SOURCE	Moraxella catarrhalis		
ORGANISM	Moraxella catarrhalis		
	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;		
	Moraxella.		
REFERENCE	1 (bases 1 to 5976)		
AUTHORS	Loosmore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.		
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella		
JOURNAL	Patent: WO 0107619-A 2 01-FEB-2001;		
	CONNAUGHT LABORATORIES LIMITED (CA)		
FEATURES	Location/Qualifiers		
source	1..5976		
	/organism="Moraxella catarrhalis"		
	/db_xref="taxon:480"		
BASE COUNT	1972 a 1360 c 1355 g 1289 t		
ORIGIN			
	Query Match 64.1%; Score 3947.8; DB 10; Length 5976;		
	Best Local Similarity 80.9%; Pred. No. 0; Mismatches 1002; Indels 156; Gaps 20;		
	Matches 4905; Conservative		
QY	166	gtatcggtgcagcgtcaatggcagtgcttatgtctcaacaattactaccaag---atc	222
Db	1	GTGATCGGTGCACGCTCACTGCGCAGTGTATTGTCTCAAAAAAAGATACCAACATATC	60
QY	223	gaatttggtcaacaacaagataaaacaacagctgaaaggcagatgccctagcgacagt	282
Db	61	GCAATTGGTGAACAAACACGCAACGCGTC-----AGGCACCTGCCAAGGCGGCGT	114
QY	283	gaagcatccattgcttttgtagtcttcttaaggcaacaggtctcaagctattgctatc	342
Db	115	GATCGAGCCATTGCTATTGTTGAAATGCTAACGCACAGGCGGTCAAGCCATCGCCATC	174
QY	343	ggtagtgtcaaacaccagatccctataatggttagtaatggttaagtgttccacgcaaa	402
Db	175	GGTAGTAGTAATAAAGTCTCAATGGAAGCAGATTGGATGAAGATAGGTACCGATGTACG	234
QY	403	ggtaacagtgctccatcgccatcggtgtagtgatttggtcagggtgatgctcgtcattgcc	462
Db	235	GGTCAAGAGTCCATCGGCATCGGTGGTGGTATGATGAAGGCTAGTGGTGATGCTCGATTGCC	294
QY	463	atcggtagtgtgacttatatttgcctaaagaatcttgcattgcattgaagaa-----	509
Db	295	ATCGGTAGTAGTACTTACATTTGCTTGTGATCAGCATGGTATCTCAACATCCGAAAGGT	354
QY	510	--tgaatttcaaaacttatcatcggccatgaaataataaaaaatacaaacctcaacc	567
Db	355	ACTCTGTATTAACGATCTTATTAAACGGCCCTGCAGTATTAAAGAATACGAAGCTCAAG	414
QY	568	gatggtaaatacaataatcgccacacaagagacaaaggcgccagtcagtctgagtgga	627
Db	415	GATATGATGTAAATAATAGACGCACAACCGCAAGCGGCACACGCGAGTACTGCGAGTGGGA	474
QY	628	gcatgtcataatgcacaggggtcattttccaacgccttggatcatcagcacaacagctgaa	687
Db	475	GGCATGTCTATATGCACAGGGTCAATTTTTCACACGGCTTTGGTACACGGCAACAGCTAAA	534

Qy	588	gctgcctattccttggcagtaggtcttggccgcccgaagccacaaacaaatcttcaatcgct	747
Db	535	AGTGCCTATTTCCTTGGCAGCTGGGTCTTGGCGCCACAGCCGAGGCCCAATCTTACAAATCGCT	594
Qy	748	gttggtttccaatgcgaagctaaacgcgttttcgagcgacagccattggtggaaatactctga	807
Db	595	ATTGGTTCTGATGCACATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGACGGTACTCGT	654
Qy	808	gttaatttgggtgcgagcgcttgccttagtttgggtcttcagatccttgataggataat	867
Db	655	GCTCAGCTACAGGGCAGTAGTTCGCCTAGTGCAAGGTTCTGTGTGTCATCTCAGAGTGATAAT	714
Qy	868	aatacagatccagtgccctatgtaccactaggtaaaaagtttagcagaccagtatataaagcc	927
Db	715	AATTCT---TAGACGGGCTTATACCAAATACCCAGGCACCTAGACCCCAAGTTTCAAGCC	771
Qy	928	accgcgcaggtgatcttcagcatattttccattggtgaatagtaataataataatagc	987
Db	772	AC---CAATAATACGAAGCGGGCTCCTTTCCATTGG-----TAGTTAAC	813
Qy	988	agtatcaggcgtaaatatcatatcgaatcggtgcgggtcttcggtgatacogcatgcggtcaat	1047
Db	814	TCATCAACAGTAAATCATCAATTCGCTGCAGGTGTTAATAAAACGATGCGGTCAAT	873
Qy	1048	gtggcacagcttaaatgttgaggaaactggcta---atcgtaaaattacttttaaaggtt	1104
Db	874	GTGGCACAGCTAGAACGGTGGTGAAGTGGGCTAAGGAGCGTGATAAAACGATGCGGTCAAT	933
Qy	1105	gatgtgacaataatagcaatagctagagaagagtttgggcaatacttttaactattaaa	1164
Db	934	GATG-----ATACAGTAGTCTGACGTAAATAATAGGTTTGGGATAATACTTTAACTATTAAA	987
Qy	1165	ggtgatcacagaccacgcattaaaccgaagctaacaatcggtgtggttaa-----cagat	1218
Db	988	GGTGGTCAGAGACCAACAGCATTAACCGATAATAATATCGGTGGTAAAGAGCGCTGAT	1047
Qy	1219	ggcaatgctgaaagttaaaactgtcctaaagctgactggat[------g	1263
Db	1048	AATAGTGGTCTGAAGGTTAACTTGTCTAAACCTTTAAACAATTTACVTACVTGAGGTGAATACA	1107
Qy	1264	accagttctccgtctaccacaaaatacccggttagtaataccaacaacaacacgcgcag	1323
Db	1108	ACTACATTAATGCCACACCACAGTTANGTGTAGTACTAGTAGTACTTACAGCTGAA	1167
Qy	1324	ctacaagcgggtgtgttgaccttagcc-----caatacaggtacaaaacacagataaa	1377
Db	1168	TTATTGTAGTGATAGTTTAACTCTTACCCAGCCCAATACAGCGCTCAAGACCAACGCAAA	1227
Qy	1378	accgtctacagcatgatggattgaagtttactaatgatatagatatagatatagaaactaaa	1437
Db	1228	ACCGTCTATGGCGTTAATGGGTGAAGTTTACTTAATAATGCAGAAACAAACAGCAGCAATC	1387
Qy	1438	ggttactactcgtattaccaaaaaagaataatggtttgcgttgactaaatgattggagttgat	1497
Db	1288	GGCACTACTCGTATTACCAGAGATAAAATTGCCTTGTCTCGAGATGGGTGATGTTGATGAA	1347
Qy	1498	gaagcaaaccttatcttgacaacgaaaagctaaagtgtgcaacagcacccctaaacagt	1557
Db	1348	AAACAGCACCATATTGTGATATAAAACAACTTAAAGTGGGTAGTGTGCAATTTACCATA	1407
Qy	1558	ggtagcttgactgttaataaacaccactggtaataacaaaalccaagtcggtgctaattggc	1617
Db	1408	GACAAATGGCAATTGATGCAGGTAAATAAAGATTCAGTTAATCTTGCCCAAGGTGAGCAGTGCT	1467
Qy	1618	attaatttggccacagtcgctaaataatgttgcaataacctcagcaacagtcggtcactgct	1677
Db	1468	AACGATGCGGTATACCATCGAACAGCTCAAGCGCCCAAGCCCTACTTTTAAACGACGCGCT	1527
Qy	1678	cytattaccgaagaaaaattggtttctggtgactaatatgatgtgagttgatagaacaagca	1737
Db	1528	GGCACTAGTGTCAACCTACTGAATATCAGTTGATGTGCTAAGAGTGGCAATGTTACCGCC	1587

Db	2617	AACGCCAAAGACATCGCCGAAAATCTAAACACCTAGCCACGAAATTCACACCAACAAA	2676
Qy	2669	ggcagcgaacacccgcccctacaaaccttaccgtttaaaaggttagatgaataatgaag	2928
Db	2677	GGCACACAGACACCGCCCTACAAACCTTTACCGCTTTAAAGGTAGATGAAATATAT	2736
Qy	2929	gctgatgacccaacgcataccgctgggtgaagatggcacaaagtggtaagtcaacacc	2988
Db	2737	GCTGATGACGCACACGCGATCACCGTGGGTCAAAAGACGCAAAATTAATCAAGTCACACCC	2796
Qy	2989	ttaaactcaaggtaaaaacgcttctgatattaaacgcgacaaagatggtacggtacc	3048
Db	2797	CTAACACTCAAGGTGAAACGGCTTTAATATTAAACCCGACAAAAATGGTAGCGTTAC	2856
Qy	3049	tttggtattacaccccaaaagcggctcttaagccggtggacagaccactctaaacaacat	3108
Db	2857	TTTGGCATTAACACACACGCGCGCTTTAAAGCGCGCAAAAGCACCC---CTAAACGACGGT	2913
Qy	3109	ggctgtctattaaaaacacccgctagtaagcaacaaacccaagtcggtgctgctggtg	3168
Db	2914	GGCTTGTCTATTAAACCCACCTGGTAGCAACAAATCCAAGTCGGTGTGATGGCGTG	2973
Qy	3169	aagtttgccatggt---taataagtggtgttagtgctggtgctggtgctggtgctggtg	3225
Db	2974	AAGTTTGGCAAGGTTAATAATAAATGTTGTAGTGTGCTGGCATTTGATGGCACAACTCGC	3033
Qy	3226	attaccagagatgaattggcttactgggactaatggctcaacttgataaaagcaaaccc	3285
Db	3034	ATTACCAGAGATGAATAATTGGCTTTACTGGGACTAATGGCTCACTTGTATAAAGCAAAACCC	3093
Qy	3286	cacctaagcaagaagcgtattaacgcaggtgggttaaaagattaccaactcaatcaggt	3345
Db	3094	CACCTAAGCAAGACGGCATTAACGACGCTGTTAAAGATTTACCAATTTCAATCAGGT	3153
Qy	3346	gagattgcacaaacagccatgatgctgtgacagcggcggcgaagatttatgtattaaaaacc	3405
Db	3154	GAGATTGCCCAAAACAGCCATGATGCTGTGACAGCGCGCAAGATTATGATTTAAACACC	3213
Qy	3406	gaactgaaataataatcagcagcaactgcacaaacagacacaaactcattacagaattc	3465
Db	3214	GAACTTGAAACAAAAATCAGCAGTACTGGCCAAACAGCACAAATCATTTACCGAATTC	3273
Qy	3466	tcagtagcagatgaacaaaggaataactttacggttagtaacccctactccagttatgac	3525
Db	3274	TCAGTAGCAGATGAACAAGGTANTAACTTTACGGTTAGTAACCTTTACTTCCAGTTTATGAC	3333
Qy	3526	acctcaagaacctctgctgctacacactttgcaggtgaaacgcgcatcaccaccaaggt	3585
Db	3334	ACCTCAAGACCTCTGATGTATCATCCTTTGACAGGTGAAACCGCATTAACCAAGGTA	3393
Qy	3586	aataaaggt	3645
Db	3394	ATAAAGGTGTGGTGGGTGGGATTTGACCAACCAAGGCCTTAACACGCGCTTAAGCTG	3453
Qy	3646	accgtgggtgaataataatggcaaggcatgtcatataacgccaataatgggtcaaaatcc	3705
Db	3454	ACCGTGGGTGAATAATAATGGCAAGGATGTCTATTGACAGCCCAAAATGGTCAAAATACC	3513
Qy	3706	atcacagactaagcaacactctagctaaattgtaccatgataaaggtagctgacgacc	3765
Db	3514	ATCACAGGACTTAAGCAACACTCTAGCTAATGTTTACCAATGATAAAGGTAGCGTACGCAOC	3573
Qy	3766	acagaaacggggcaataataataaagacgaacaaacccgctgcgcagcattgttgat	3825
Db	3574	ACAGAACAGGGCAATATATCAAGACGCAAGACAAACCCCTGCGCCAGCATTTGTGAT	3633
Qy	3826	gtgctaagcgcaggttttaacttgcgaagcaatggtgaagcgttgcaggttgcctcaact	3885
Db	3634	GTGCTAAGCGCAGGCTTTAACTTTGCAAGGCAATGGTGAAGCGGTTGACTTTGTCTCCACT	3693
Qy	3886	tatgacacccgtcaactttgcgaatggcaataaccacccgcgtgaaggtgaacctatgtac	3945
Db	3694	TATGACACCGCTCAACTTTGCCGATGGCAATGCCACACCGCTTAAGGTGACCTATGATGAC	3753
Qy	3946	acaagcaaaacccagtaaaagtgtctctatgatgtaagtggatgatacaacaccattgaagtt	4005
Db	3754	ACAAGCAAAACCCAGTAAAGTGGTCTATGATGTCATTTGGATGATGATACNACCATTGAAGTT	3813
Qy	4006	aaagataaaaaaacttggtgtaaaaacacacattgaccaggtactggtgacaggtgctaat	4065
Db	3814	AAAGTAAAAAAACTTGGCGTAAAAACCCACCATTTGACCAGTACTGGCACAGGTGCTAAT	3873
Qy	4066	aaatttgccctaagcaaatcaagctactggtgcatgagcctgttcaagccagtgatcgtt	4125
Db	3874	AAATTTGCCCTTAAGCAATCAAGCTACTTGGCGATGGCGCTTGTCAAGGCCAGTATCGTT	3933
Qy	4126	gctcatctaaacaccttatctggcgacatccaaactgccaaggggcaagcgaagcgaac	4185
Db	3934	GCTCATCTTAACACACTTATCTGGCGACATCCAAACTTGCCAAAGGGCAAGCCAAAGCAAC	3993
Qy	4186	aactcagcaggtctgttggtatggtgcaataaggtcatctatcacagtcacgataac	4245
Db	3994	AACTCAGCAGGCTATGTGGATGCTGATGGCAATAAGGTGATCTATGACAGTACCGATAC	4053
Qy	4246	aagtactatcaagcaaaaaatgagcacagttgataaaacccaagaagtgtgccaagac	4305
Db	4054	AAGTACTATCAAGCCAAAAATGATGGCACAGTTGATAAAAACCAAGAAAGTTGCCAAGAC	4113
Qy	4306	aaactggtgcgccaagcccaaacccccagatgggacatttggctcaaaatgaatgcaaatca	4365
Db	4114	AAACTGGTGGCCCAAGCCCAACCCAGATGGCACATTTGGCTCAAAATGATGTCNAATCA	4173
Qy	4366	gctattcaaaagaacaagtaaatgataatgccaataaaagcgaaggtatcaatgaagacaac	4425
Db	4174	GTCAATTAAACAAAGAACAGTAATGATGCCAATTAAGCAAGCAAGCATCAATTAAGAGACAC	4233
Qy	4426	gcctttgttaaaggacttgaaaaagcccttctgtataacaaaccccaaaacccgcgcagta	4485
Db	4234	GCCTTTGTTTAAAGGACTTGAAGAACCGCTTCTGTATACAAACCCCAAAACCGCCGACGTA	4293
Qy	4486	actgtgggtgatttaaatccgttgcacaaacacccgctgaccttgcagggggatatacagcc	4545
Db	4294	ACTGTGGGTGATTTAANTGCCGTGGCCAAACACACCGCTGACCTTTTGACAGGGGATACAGCC	4353
Qy	4546	acaacggttaaaaaacccggtggagacttgcacataaaggtggcgaacacacacccaat	4605
Db	4354	ACAACGGCTAAAAAATCTTAAACAGCGTTAATGACAGGTGATGACCATCAAGAGTGGCCAAAC	4413
Qy	4606	aagctaacccgataataacatcggtgtgtgtgtagcaggttactgtagtgcacttgcacaaactt	4665
Db	4414	AAGCTAACCCGATTAATACATCGCTGTGGTAGCAGGTACTGATGGCTTCACTGTCAAACTT	4473
Qy	4666	gccaaagaccttaacaaactttaaagcaggttaataagcaggtggcgaacaaatgtatgaataa	4725
Db	4474	GCCAAAGACCTTAACCAATCTTAACAGCGTTAATGACAGGTGATGACAGGTGATGACCAAA	4533
Qy	4726	ggcactctttgttagcgcgaacgcgttcaagcgaagcaaacacccctgtgtgaagtgcc	4785
Db	4534	GGCTGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAAAACACCCCTGTGTTAAGTGCC</	

QY 689 ctgctattcttgccagtagcttctgcccgaagcacaacaaactcttcaatcgtg 748
 Db 1589 CAACCTTACAAACA -----TTGGCGTGAACACACCCAGAGCTTAACAGTGTGCACTA 1639
 QY 1799 ttaatgctggttaatacacaagattaccggacttactaaatggtatagcaaaataccgagtcg 1858
 Db 1640 GTGATAAATTTAGTGTAAAGGGTAGTGTACGAACAATAGCTTACCTCCGCGCAACATTT 1699
 QY 1859 ttacacaaacagctcacaagagcgaagcctactttaaacgagcgatggcactcagta 1918
 Db 1700 TGGCAAGCTATCTTAATGAAGTCAATCGAACGCTGACAGTCTCTACAAAGCTTTACCG 1759
 QY 1919 ttaatgtagtaacgggagcttagttgtagtagtggcgaatattaccaccccaactata 1978
 Db 1760 TTAAGAAGAACGATGATGACG-----CCAACG 1789
 QY 1979 acattgagtgtaaaaaaccacaaagcttaacagtaagtgccaccagtggttaataataattta 2038
 Db 1790 CTATCACCGGTGGCTAAAGATACGACAAAANATGCCGCGCAGTCAGCACTTAAACTCA 1849
 QY 2039 gtgttagtaatgctcatgatacaaatagcttagttaccgcccgaagatttggcagactatc 2098
 Db 1850 AAGTAAACCGGTCTAACGGTTGCTACCAAAAAGATGGTAGCGTTACCTTTGGGCTTA 1909
 QY 2099 taaataagtcgaatgaacggctgacagtgctctcaaacgtttaaagtcacaaacgggtg 2158
 Db 1910 GCCAAGATAGCGTCTGACCATTTGGCAAAAGACCCCTTAACACAGTGGCTTGACTGTGA 1969
 QY 2159 ataataagcaaacgcatcaccgtgggttaaaagatacaaacggaagacccctcaaacct 2218
 Db 1970 AAGATACCAACGACAACTCAACCTCGGTCTTAATGGCACTTAAATTTACCTTAATGTGA 2029
 QY 2219 taaactcaaacggtgaacgggtgttaattacgacaaatagagccacaggtacagttta 2278
 Db 2030 GTAGTAATCCAGGTACTGGCATTTGCNAATACCGCTCGCATTA---CCAGAGATAAATTTG 2086
 QY 2279 ccttggcattgacaaagtaagtgctcaccacgcttaagctgacgggtgggtagcgata 2338
 Db 2087 GCTTTGTGCTGTGTGTTGGTGCAGTTGTATC-----AAACAAACCTTATCTTGATCAA 2139
 QY 2339 caaatggtaactgattggttattgagcaagtccttagcgtgacgggtgaacagaccccaaa 2398
 Db 2140 GACAAGCTA--CAAGTTGGCAATGTTAAGATTACCAACACTGGCATTAACCCAGGTGGTA 2197
 QY 2399 acatcattaaagattgtcccaacactgcttagcattgcccagtcacagtgccgcgaaca 2458
 Db 2198 AAGCCATCAGAGGCTGTCCCAACACTGCTAGCATTTGCCGATCAAAAGTAGCGCAACA 2257
 QY 2459 tagcactgggaatacaatcgaagaaaaagcaaatcccaacgctgcccaggaattgatgatg 2518
 Db 2258 TAGAACTGGGCAATACAATCCAAAGACAAAGCAATATCCAAACGCTGCCAGCAATTAATGATA 2317
 QY 2519 tgcataatgcaagctttaaacttaaaataatggaagacaaagactttgtctccactt 2578
 Db 2318 TATTAATACAGGCTTTAACTTAAATAAATAATAACACCCCAATTTGACTTTGTCTCCACTT 2377
 QY 2579 atgacactgttgaactttatcgatggcaatggccacccacgcccacagtaacttatgatgaag 2638
 Db 2378 ATGACATTGTGACTTTTGCAATGGCAATGGCAATGGCAATGGCAATGGCAATGGCAATGG 2437
 QY 2639 coaatcaaacccagtaaaagtgccgtatgtagtgaatgtagatgagaacacattgaactga 2698
 Db 2438 CTAAACAAAACAGTAAGTGGTATGATGTGAATGTGAATGTGAATGTGAATGTGAATGTGA 2497
 QY 2699 caggcgataatggcgaagaaacacttggcggtcaaaacacacacacacacacacacacacac 2758
 Db 2498 CAGGCACTGATGACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2557
 QY 2759 ctaatggtaaatgcaac-----tacatttagtaccagacagatgacattgcttcttga 2809
 Db 2558 CTAAATGGTAATACAGCAACTTAATTTAATGTGTAACTCTAGTGTGAATGTGAATGTGAAT 2617
 QY 2810 aagccagtgatagtcgcccggcaatctcaaacaccccttagccgaggaatttcacacacacac 2869

QY	5018	ataacgataccgaaataacttccactgtgtgtgtacagtggtggtgataaaagacggca	5077	QY	6098	tcaatggttcagccgatacccaagcccatgtaggggcgagcttggtgacaggttttccact	6157
Db	4835	ATAACGATACCGAAATCTTCCACTGTGTGTATACAGTGGGGTGTGATAAAGACGGCA	4894	Db	5915	TCATATGTTACGCCGATACCCAAAGCCCATGTAGGGCGGCAGTTGGTGCAGGTTTTCAT	5974
QY	5078	acgttaacggcgatttaagcaatgtttgggtcaaaaccccaaaaagatggcagcaaaaaag	5137	QY	6158	tt 6159	
Db	4895	AGCTAACGGCGATTAAAGCAATGTTGGGTCAAAACCCAAAAAGATGGCAGCAAAAAG	4954	Db	5975	TT 5976	
QY	5138	ccctgctccacttatcaacgcgcgagtcagacaactatgtgaccaaaaccccccgag	5197	RESULT	9		
Db	4955	CCCTGCTCCCACTTATACAGCGCGAGGTACAGCAACTATTTCACCAACAACCCCGCAG	5014	AX079957	AX079957	3135 bp	DNA
QY	5198	aagccattacagaaataaatgacaaggtatccgcttcttcattcgaatgcaacgatgcgaac	5257	LOCUS	Sequence 45 from Patent WO0107619.	PAT	22-FEB-2001
Db	5015	AAGCCATTACAGATAAATGAACAAGGTATCCGCTTCTTCATGTCAACGATGGCAATC	5074	DEFINITION	AX079957		
QY	5258	aagcctgtgttacaagggtaacggcattgactcaagtcctcaggcgaagcaactcag	5317	ACCESSION	AX079957		
Db	5075	AAGAGCCTGTGTACAAAGGCGTAAAGCGCATTTGACTCAAGTCCCTCAGCAAGCACTCAG	5134	VERSION	AX079957.1	GI:13159465	
QY	5318	tgccgataggttccaggccaagcgagatgggtgaagccgcgttgcattagcgacacaaa	5377	KEYWORDS			
Db	5135	TGCGCATAGGTTCCAGGCCAAGCAGATGGTCAAGCCCGCTTGCCATAGGACAGACAAA	5194	SOURCE	Moraxella catarrhalis.		
QY	5378	cccaagcaggcaacaaatccatccgcatcggtgataacacacacacacacacacacacac	5437	ORGANISM	Moraxella catarrhalis		
Db	5195	CCCAAGCAGGCAACCAATCATCGCCATCGGTGATACGCACACACACACACACACACAC	5254	REFERENCE	1 (bases 1 to 3135)		
QY	5438	ccatcgcatcggtacagggcaatgtgtgtagcaggtgaagcaactctgtgcatcgcgacc	5497	AUTHORS	Loosmore, S.M., Sasaki, K., Yang, Y.P. and Klein, M.H.		
Db	5255	CCATCGCCATCGGTACAGGCAATGTGTGTACAGGTAAAGCACTGTGTGCGCATCGCGACC	5314	TITLE	Recombinant high molecular weight major outer membrane protein of moraxella		
QY	5498	caagcactgttaagcctgataaacgttacagttacagttggtggttaataaacaacacacac	5557	JOURNAL	Patent: WO 0107619-A 45 01-FEB-2001;		
Db	5315	CAAGCACTGTTAAGGCTGATAAACAGTTACAGTGTGGGTAAATAACACACACACACAC	5374	CONNAUGHT LABORATORIES LIMITED (CA)			
QY	5558	ccactcaaacgatgtcttctgtgtgggcaataacatcacctgacagcgaagtaactcgg	5617	FEATURES	Location/Qualifiers		
Db	5375	CACTCAAACCGATGCTTGTGTGGCAATACATCACCGTGACCGTAACGTAACGTCGG	5434	source	1. 3135		
QY	5618	ttgcttaggttcaaacctctccatcagtcagtcagtcagtcagtcagtcagtcagtcagtc	5677	organism	"Moraxella catarrhalis"		
Db	5435	TTGGCTTAGGTCAAACTCTGCCATCAGTCAGTGAGGTCAGCAACCGGTACCGTTAAAGGCT	5494	db_xref	"taxon:480"		
QY	5678	aatctgacggcacacagcaggtacacacacacacacacacacacacacacacacacacac	5737	BASE COUNT	1011 a 746 c 741 g 637 t		
Db	5495	AATCTGACGGCACACAGGTACACACACACACACACACACACACACACACACACACAC	5554	ORIGIN			
QY	5738	ttgcttggaacaaacggcgttggtggtggtggtggtggtggtggtggtggtggtggtggtg	5797	Query Match	49.28;	Score 3031.8;	DB 10; Length 3135;
Db	5555	TTGCTGGACAAACCGCGGTGTGGTGGGTGTGGTGGGTGTGGTGGGTGTGGTGGGTGTGG	5614	Best Local Similarity	98.78;	Pred. No. 0;	
QY	5798	tccaaatgtgacagcaggtgaggtcagtgccacacacacacacacacacacacacacacac	5857	Matches	3093;	Conservative	
Db	5615	TCCAAATGTGCGCAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG	5674				
QY	5858	agttgtacaagccaccccaaacgattgcaaacgcaacacacacacacacacacacacacac	5917				
Db	5675	AGTTGTACAAAGCCACCCCAAGCATTTGCCAACGCAACCAATGAGCTTGACCACTGATACC	5734				
QY	5918	accaaacgaaataaagcgaatgcagggtatttcacagcgatggcgatggcgatggcgatggc	5977				
Db	5735	ACCAAAACGAAATATAGGCCAATGCAGGGATTTCATCAGCATGCGATGCGGTGCGCATGC	5794				
QY	5978	cacaagcctacattccttgccagatccatgttaccgggggtattgccaccacaacgcgtc	6037				
Db	5795	CACAAGCCTACATTCCTTGGCAGATTCATGGTTACCGGGGGGTATTTGCCACCCCAACAGG	5854				
QY	6038	aagtcggtgagcagtgaggactcgaaagctgctgataatggtcgaatgggtatttataaa	6097				
Db	5855	AAGGTGGCGTGGCAGTGGGACTGTCAAGAGCTGTCCGGAATATGTCATATGGTCAATGGT	5914				

Qy	3382	ggcaagatttatgattttaaaacccgaacttggaaataaaatcagcagctactgcgcaaaaca	3441
Dy	3384	GGTCGGGTttATGATttTAAAAACCGAACTTtGAAGCAAAATCAACAGTGTGTCTAAACA	3423
Qy	3442	gcacaaactcattacacgaattctcagtcagcagatgaacaaagtaataactttaoqgtt	3501
Dy	3444	GCACAAACTcATttACACGAATtTCTCAGTAGCAGATCAACAAGTAATCACTTTACGGTT	3303
Qy	3502	agtaaccttactccagttatgacacctcaaaagacctctgatgtcatcacctttgcaggt	3561
Dy	3304	AGTAACCGTtTACTCCAGTtTATGACACCTCAAGACCTTCTGATGTCACTCTTTCAGGT	3363
Qy	3562	gaaacggcattaccaccaaggttaataaaagttggtgctggtggcattgaccaaacc	3621
Dy	3364	GAAAACGGCATtTACCACAGTAAATTAAGGTGTGGTGGCTGTGGCAATTGACCAAAACC	3423
Qy	3622	aaagctttaaccacgccttaagctgacctgggttaataataatggcgaagcattgtcatt	3681
Dy	3424	AAAGCTtTAACACGCTTAACTGACCTGCGGTAAATATAATGCAAGGCATtTGTCTAT	3483
Qy	3682	aacgccaataatggtcaaaaataccatcacaggactaagcaacctctagctaaatgttacc	3741
Dy	3484	GACAGTAAGATGCTCAAAATACCATCACAGACTAAGCAACACTCTAGCTAAATGTTTACC	3543
Qy	3742	aatgataaaggtagcgtcgcaccacagaaacagggcaataataataaaagacgaagacaaa	3801
Dy	3544	AAATGA-----TGGTGCAGGACACGCATAAGCCAAAGGGCTTGGCAATGACACCGACAAA	3597
Qy	3802	accctgcgcgcagcattgttgatgtgtaagcgcagcctttaacttgcaggcaatggt	3861
Dy	3598	ACCCGTGGCCGACGATtTGGTATGTGTCTAAACGACGCTTTAACTTGCAGGGCAATGGT	3657
Qy	3862	gaacgggttgactttgctccacttatgacacgctcaactttgcaatggcaataccacc	3921
Dy	3658	GAAGCGGTtTGACTTTGTCTCCACTtTATGACACTGTTGACTTTATCGATGGCAATGCCACC	3717
Qy	3922	accgctaaagtgacctatgatgacacaaagcaaaacaggtaaagttgggtctatgtctaat	3981
Dy	3718	ACCGCTAAGGTGACCTTATGATGACACACAGCAAAACAGCTAAAGTGGTCTATGATGTCAAT	3777
Qy	3982	gtgatgatcaacacattgaagttaaa---gataaaaaacttggcgttaaaaaccaccaca	4038
Dy	3778	GTGGATTAATTAACACCATtTGAAGTGACACAGTGTATAAAACTTGGCTCAAAACCACCACA	3837
Qy	4039	ttgaccagttactggcacaggtgctataaatttgccttaagcaataaagctactggcgat	4098
Dy	3838	CTGACCAAAACAAAGTCTTAATGGTAATGCA---ACCAAAATtTATGCCCGCATGGCGAT	3894
Qy	4099	ggcgtgtcaagcgaagtatcgtgtctatctaataacaccttacttggcgacatccaa	4158
Dy	3895	GCCTTGTtTAAAGCCAGTGTATCGCCACCATCTTAATACCTTGGCTGGCGACATCCAA	3954
Qy	4159	actgcaaaaggggcaagcgaaggaacaactcagcaggctatgtggatgctgatgcaat	4218
Dy	3955	ACGCCCAAAAGGGCAAGCCAAAGCAGCTCAGCAAGCTATGTGGATGTGTATGGCAAC	4014
Qy	4219	aaggtcatctatgacagttaccgatacaagctactataagccaaaaatgatggcacgtt	4278
Dy	4015	AAAGTCACTCTATGACACGTACCGATAAGAAAGTACTATCAAGTCAATGACAAGGTCTAAGTG	4074
Qy	4279	gataaaaccaaaagttgcacaaagcaaaactggtgcgaagcccaaaccccgatggc	4338
Dy	4075	GACANAAACAAAGATtTGCAANAGACAAACTGGTCGCCCAAGCCCAACCCCGATGGC	4134
Qy	4339	acattggctcaaatgaaatgcgaatctaatacaaaagcaacagtaaatgatgccaat	4398
Dy	4135	CAATtTGGCTCAAAATGAATGTCAAAATCACTAGTCTATtTAAACAAGAGCAAGTAAATGATGCCAAT	4194
Qy	4399	aaaaagcaaggcatcaatgaagacaacgcctttgttaaaggacttgaaaagcgcgtctct	4458
Dy	4195	AAAAAGCAAGGCATCAATGAAGACACACGCCCTtTATCAAAAGGGCTTGAACGCCGCCAAA	4254

Qy	4459	gatacaaaacccaaaaacgcgcagtaactctgggtgatttaaatgcgttgcccaaacca	4518
Db	4255	GACACCAAAACCAAAAACGCCGAGTAACTCTGGTGATTAAATGCCGTGCCCCAACA	4314
Qy	4519	ccgctgacctttgcaggggatacaggcacacacgcgttaaaaaactggcgagactttgacc	4578
Db	4315	CCGCTGACCTTTGTCAGGGGATACAGGCACAAACGGCTAAAAAACTGGCGGAGACTTTGACC	4374
Qy	4579	atcaaaagttgggcaaacagacacacaaataagctaaccgataataaacatcgggtggtgagca	4638
Db	4375	ATCAAAAGTGGGCAAAACAGACACACCAATTAAGCTAACCGGATATTAACATCTGGTGTTGATGCA	4434
Qy	4639	ggtactgattgcttcactgtcaaaacttgccaaagacctaaccatactttaacagcgttaatt	4698
Db	4435	GGTACTGATGGCTTCACTGTCAAACTTGCCAAAGACCTAACCAATCTTAACAGCGTTAAT	4494
Qy	4699	gcaggtggcaccacaaattgataaaaaaggacatctctttttagacgcacaaacggtcgaagcc	4758
Db	4495	GCAGGTGGCACCCAGAAATTTGATGA AAAAGGCATCTTTTGTAGACGCAAAACGGTTC AAGCC	4554
Qy	4759	aaagcaaacacccctgtactaagtgccaaatggcctggacctgggtggc	4806
Db	4555	AAAGCAAAACACCCCTGTGCTAAGTGCCAACTGGGTGGNACCTGGGTGGCAACGCATCAGT	4614
Qy	4807	-----	4806
Db	4615	AACATCGGTGCAGCTGTTTGATGAATAAGATGCGGTGAACCTTTAAAGCAGTTTAAATGAAGTT	4674
Qy	4807	-----	4806
Db	4675	GCCAAAACGGTCAACAACACTAAACAACAAAGTAACCTCAGGTGGCTCATTTACCCCTTTGTG	4734
Qy	4807	-----	4806
Db	4735	GTAACCGATGCCAATGGCAAGCCCATCAATGSCACCGATGGCAAGCCCCCAAAAAGCCATC	4794
Qy	4807	-----	4806
Db	4795	AAGGGCGCGATGGTAAATACTATCAGCCAAAGCCCAACGGCGTAACCTGTGGACAAAGAT	4854
Qy	4807	-----	4806
Db	4855	GGCAAGCCCATCACCGATCGGACAAACTTGCCAACTCTGGCAGCTCATGGCAAAACCCCTT	4914
Qy	4807	-----	4806
Db	4915	GATGCAGGTCAATCAAGTGTGGCAAGCCTTAGCGGCAACTAGATGCCATCACCCCTAACC	4974
Qy	4807	-----	4806
Db	4975	AACATCAAGTCCACTTTGCCACAAATTGACACACCAACACACAGGTAATGCCAATGCAGGG	5034
Qy	4807	-----	4806
Db	5035	CAAGCCCAAAAGTCTGCCAGCCCTATCAGCAGCACAGCAAAAGTAATGCTGCCAGTGTCAAA	5094
Qy	4807	-----	4806
Db	5095	GATGTGCTAAATGATAGCGTTTAACCTTGACAGCAACCAATCACAATCAAGTGGACTTTGTCAAA	5154
Qy	4807	-----	4806
Db	5155	GCCTATGATACCGTCAACTTTTGTCAATGGTACAGGTGCCGACATCAACAGCGTGCCTAGT	5214
Qy	4807	-----	4806
Db	5215	GCTGATGGCAGGATGAGTAACATCACCGTCAACACCGCCCTTACAGCGGACCGATGATGAT	5274
Qy	4807	-----	4806
Db	5275	GGCAATGTGCTTATCAAAAGCCAAAGATGGTAAAGTTCTTACAAAGACAGCACCTCATGCCCA	5334
Qy	4807	-----	4806

Db 5335 AACGGCTCACTAAAGCAGGCAAAATCAGCCAGTGTATGCCAAAACCTCCAACTGGTCTAAGC 5394
QY 4807 ----- 4806
Db 5395 CTTGTTAAACCCCAATGCTGGTAAAGGCAGTACAGGCGATGCAAGTGGCTCTTAAATAACTTA 5454
QY 4807 ----- 4806
Db 5455 TCAAAACGGGTATTTAAATCCAAAAGATGGTACAACTACTACCAAGTAAAGCTCTGATGGC 5514
QY 4807 ----- 4806
Db 5515 ATCAGTATCCNAGCAAGATAACAGCAGCATCACCTTACCAAGATGGCTGAATGTA 5574
QY 4807 ----- aagtcatacagtaaatgtgggcaaaagggcacaagaataaccgcgctgccaatgta 4860
Db 5575 GCGGTAAGGTCATCAGCAATGTGGTAAAGGCACAAAAGACACCGAGCTGCGCAATGTA 5634
QY 4861 caacagttaaacgaagtacgcaacttgttgggtct-----tgtaatagataacgct 4911
Db 5635 CAACAGTTAAACGAAGTACGCAACTTGTGGGTCTTGGTAAATGCTGTTAATGATAACGCT 5694
QY 4912 gacggcaatcaggttaacattgcgcacatcaaaaagaccccaaatcaggttcacatct 4971
Db 5695 GACGGCAATCAGGTAAACATTGCCGACATCAAAAAGACCCAAATTCAGGTTTCATCT 5754
QY 4972 aaccgcactgtcataaagcagggcagcgtactgtgcgttaaggttaataacgataccgaa 5031
Db 5755 AACCGCACTGTATCAAAGCAGGCACGCTACTTGGCGTAAAGGTAAATTAACGATACCGAA 5814
QY 5032 aaacttgccactggtgtgtacaaagtggcggtgataaaaagcgcacagctaaacggcgt 5091
Db 5815 AAATTCGCCACTGTGTGTGTACAAAGTGGCGGTGGATAAAGACGGCAACGCTTAACGGCGAT 5874
QY 5092 ttaagcaatgttgggtcaaaaccccaaaagatgagcagcaaaaagccctgtcgcact 5151
Db 5875 TTAGCAATGTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCCACT 5934
QY 5152 tataccgcgaggtcagacaaactatgtgaccaaaccaccccgagagccattgacaga 5211
Db 5935 TATAACGCGCAGGTGACAGCAACTATTGTACCAACAACCCCGCAGAGCCATTGACAGA 5994
QY 5212 ataatgaacaaggtatccgtcttccatgtcaacgatgggcaataaagacgttgtgta 5271
Db 5995 ATAAATGAACAAGATATCCCGTCTTCATGTCAACATGGCAATCAAGAGCCCTGTGSTA 6054
QY 5272 caagggcgtaacgcattgactcaagtgccctcagcagcaactcagtgccgataggttc 5331
Db 6055 CAAGGGCTAACGCGATTGACTCAAGTGCCCTCAGGCAAGCACTCAGTGGCGATAGGTTTC 6114
QY 5332 caggccaagcgagatggtgaagccgcgttggccatagggcagacaaacccaagcagggcaac 5391
Db 6115 CAGGCCAAGCGAGATGTTGAAGCGCGGCTTGCCATAGGCAGACAAACCAAGCAGGCAAC 6174
QY 5392 caatccatgccatcggtgataaacgacaagcaagcagggcgatcaatccatcgcatcgtt 5451
Db 6175 CAATCCATCGCCATCGGTGTAAAGCAAGCAAGCAGCGGCGATCAATCCATCGCCATCGGT 6234
QY 5452 acagggcaatgtgtagcaggttaagcactctgtgtccatcgccgagcccaagcactgttaag 5511
Db 6235 ACAGGCAATGTGTTAACAGTTAAGCACTCTGGTGGCTATCGCGACCCCAAGCACTGTTAAG 6294
QY 5512 gctgataacagttacagttggtggttaatacaaacaggtttaccgatgccaactcaaacgcat 5571
Db 6295 GCTGATAACAGTTACAGTGTGGTAAATAACAACCAAGTTTATCGATGCCACTCAGACCGAT 6354
QY 5572 gtctttggtgtgggcaataacatacccggtgaccgaaagtaactcggttgccttaggttca 5631
Db 6355 GTCTTTGTTGGTGGCAATAACATCACCGTGTACCGGAAAGTAACTCGGTGCTTAGGTTCA 6414
QY 5632 aactctccatcagtcagggcacacacgagggcacacaaagcccaaaatctgacggcaca 5691
|||||

Db 6415 AACTCTGCCATCAGTGCAGGCACACACAGCAGCACACAAAGCCAAAANAATCTGACGGCACA 6474
QY 5692 gcaggtacaacacacacagcaggtgccaaggtacaggttaaggtcttgcgtggacaacg 5751
Db 6475 GCAGTCAACACCACACACAGCAGGTGCAACAGGTACGGTTAAAGGCTTTGCTGGACAACG 6534
QY 5752 gcggttggtgggtctccgttgggtgctcaggtgctgaacgcgtatccaaaatgtgca 5811
Db 6535 CCGTTTGGTGGGTCTCCGTGGGTGCTCAGGTGCTGAACGCCGTATCCAAAATGTGCA 6594
QY 5812 gcaggtgaggtcagtgccaccagcaccgatcggtgcaatggtagcaggtgtacaaagcc 5871
Db 6595 GCAGGTGAGGTCAAGTGCACACAGCAGCAGCATGGGTCAATGTTAGCCAGTTGTACAAAGCC 6654
QY 5872 acccaagcattgccaacgcaccaatgagcttgaccatcgatccaccaaaaacgaaaat 5931
Db 6655 ACCCAAGGCATTGCCAACGCAACCAATCAGCTTGACCATCGTATCCACCAAAACGAAAAT 6714
QY 5932 aaagccaatgagggatttcacagcagtgatggcgatggcgatccatcccaaacgacctacatt 5991
Db 6715 AAAGCAATGAGGATTTTCATCAGCGATGGCGATGGCGTCCATGCCACAAGCCTACATT 6774
QY 5992 cctggcagatccatggttacccgggggtattgccaccacaacggtccaaggtgcggtgca 6051
Db 6775 CTTGGCAGATCCATGTTACCGGGGTATTGCCACCACACAGGTCAAGTGGCGTGCA 6834
QY 6052 gtgggactgtcgaagctgtcggataatggtcaatgggtatttaaaatcaatgggttcagcc 6111
Db 6835 GTGGACTGTGCAAGCTGCGGATTAATGGTCAATGGGTATTAAATCAATGGTTTCAGCC 6894
QY 6112 gatacccaagccatgtaggggcgaggttggtgcaggtttttcacttt 6159
Db 6895 GATACCCAAGGCCATGTAGGGGCGGAGTTGTCAGGTTTTTCACTTT 6942

RESULT 11
AX079959
LOCUS AX079959 2448 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 47 from Patent WO0107619.
ACCESSION AX079959
VERSION AX079959.1 GI:13159466
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
REFERENCE 1 (bases 1 to 2448)
AUTHORS Loosmore, S.M., Sasaki, K., Yang, Y.P. and Klein, M.H.
TITLE Recombinant high molecular weight major outer membrane protein of moraxella
JOURNAL Patent: WO 0107619-A 47 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
Location/Qualifiers
source 1..2448
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 826 a 569 c 556 g 497 t
ORIGIN

Query Match 38.2%; Score 2352.6; DB 10; Length 2448;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2412; Conservative 0; Mismatches 24; Indels 15; Gaps 3;

QY 3037 ggtcgggttaacttggcattaaacaccccaagcgttcttaagccgagcagaccact 3096
Db 1 GGTACGGTTACTTGGCTTAAACACCAACAGCGGTCTTAAAGCCGCAAAAGACC--- 57
QY 3097 ctaacaacaatggtctgtctattaaaaaacccgctagtagtaacgaacaaatccaagtcgt 3156
Db 58 CTAAACGACGGTGGCTTCTCTATTAAAAACCCCACTGGTAGGGAACAATCCAAAGTCGGT 117
QY 3157 gctgatggcgtgaagtttgccatggt---taataatggtgtgttagtgctggaattgat 3213

Db 118 GCTGATGGCGTGAAGTTGGCCAAAGTTAATAATATGTTGTGTAGGTGCTGGCATTTGAT 177
 QY 3214 ggcacaaactcgcattaccagagatgaaatggctttactgggactaatggctcacttgat 3273
 Db 178 GGCACAACTCGCATTTACAGAGAGTGAATGGCTTTACTGGGACTAATGGCTCACTTGAT 237
 QY 3274 aaagcaaaacccacccttaagcaaaagcagcgttaacgcaggtggttaaaagatatacaac 3333
 Db 238 AAAAGCAAAACCCACCCTAAAGCAAGACGGCAATAACGAGGTGGTAAAGATTAACCAAC 297
 QY 3334 attcaatcaggtagagattccaaaacacccatgatgctgtgacagcgcgcaagatttat 3393
 Db 298 ATTCAATCAGGTGAGATTGCCAAACACCCATGATGCTGTGACAGCGCGCAAGATTAT 357
 QY 3394 gatttaaaacccaacttgaaataaaatcagcagctactgccaacacgacgacaaactca 3453
 Db 358 GATTTAACCAACCGAATTTGAAAACAAATCAGCAGTACTTGCCAAAACAGACAAAATCA 417
 QY 3454 ttacagaaattctcagtagcagatgaacaaagtaataactttacggttagtaaccccttac 3513
 Db 418 TTACACGAATCTCAGTAGCAGATGAACAGGTAAATAACTTTACGGTTTGTAGTAACCCATTAC 477
 QY 3514 tccagttatgacacctcaaaagacctctgagtcatcacctttgcaggtgaaacggcatt 3573
 Db 478 TCCAGTTATGACACCTCAAGACCTCTGATGTCATCACCTTTGCAGGTGAAAACGGCATT 537
 QY 3574 accaccaagtaataaaggtgtggtgctgtggtgattgacaaacaaaggttcaacc 3633
 Db 538 ACCACCAAGTAATAAAGGTGTGGTGTGGGATTGACCAACCAAGGCTTTAACCC 597
 QY 3634 acgcttaagctgacgtgggtgataataatggaagagcattgtcattacagcgaacaaat 3693
 Db 598 ACGCTTAAGCTGACCGTGGGTAAATAATGGCAAGGCAATTGTCAATGACAGCCAAAAT 657
 QY 3694 ggtcaaaataaccatcacaggaactaagcaacactctagcttaagttaattaccatgataaagt 3753
 Db 658 GGTCAAAATACCAATACAGACTAAGCAACACTCTAGCTAATGTTACCAATGATAAAGGT 717
 QY 3754 agctagcaccacagacaggggcaataataacaaagcaagaaagacaaacccgtgcgcc 3813
 Db 718 AGCTAGCCACACAGACAGGCGCAATATATCAAAAGCAAGCAAAACCCGTGCGCC 777
 QY 3814 agcattgttgatgtgctaaagcagggctttaacttgcagagcaaatggtgaagcggttgac 3873
 Db 778 AGCAATTGTGATGTCTAAGCGCAGGCTTTAACTTGCAGGCAATGTGTGAAGCGGTTGAC 837
 QY 3874 ttgtctccacttatgacaccgtcaacttgcgaatggcaataccacacccgcgttaagtg 3933
 Db 838 TTTGTCTCCACTTATGACACCGTCAACTTTGCCGATGGCAATGCCACCCGCTAAGGTG 897
 QY 3934 acctatgatgacacagcaaacaccagtaaaagtggtcttatgatgtcaatgttgatgataca 3993
 Db 898 ACCTATGATGACAAAGCAAAACCAAGTAAAGTGTCTATGATGTCAATGTGGATGATACA 957
 QY 3994 accattgaagttaaagataaaaaacttgcgttaaaacaccacacattgaccagttacggc 4053
 Db 958 ACCATTGAAGTTAAAGATAAAAACTTGGCGTAAAAACCCACCATTTGACCACTACTTGGC 1017
 QY 4054 acagtgcttaataaatitgccctaagcaatcaagctactggtgagctggttgcgaagcc 4113
 Db 1018 ACAGGTGCTAATAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGCGCTTGTCAAGGCC 1077
 QY 4114 agtgatactgttgcatacaaacaccttatctggcgacatccaaactgcgaagggca 4173
 Db 1078 AGTGATATGTTGTCATCTAACAACCTTATCTGGCGACATCCAAACTGCGCAAGGGGCA 1137
 QY 4174 agccaagcaaaactcagcagcgttatgttgatgctgtgagcaataaaggttcatctatgac 4233
 Db 1138 AGCCAAGCAAAACACTCAGCAGGCTATGTGGATGCTGATGGCAATAAGGTCATCTATGAC 1197
 QY 4234 agtacogatacaacagttactatcaagcaaaaatgatggcacagtttgataaaacaaagaa 4293

Db 1198 AGTACCGATTAACAGTACTATCAAGCCAAATAATGATGGCAGCTTGATTAACCAAGAA 1257
 QY 4294 gttccaaacacaaactggtcgcccaagcccaaaacccagatggcacattgggtcctcaaatg 4353
 Db 1258 GTTGCCAAACAAACTGTTGTCGCCAACAGCCAAACCCAGATGGCATTGGCTCAATG 1317
 QY 4354 aatgtcaaatcagtcattaaacaaagaacaaagtaaatgagccaaataaaagaagaagcatc 4413
 Db 1318 AATGTCAATCAGTCAATTAACAAAGAAACAAAGTAAATGATGCCAATAAAAGCAAGCATC 1377
 QY 4414 aatgaagaacaacgctctgttaaggaactgaaaaagccgctctctgatacaaaaaacaaa 4473
 Db 1378 AATGAAGACAAACGCTTGTGTTAAAGGACTTTGAAAAGCCGCTTCTGTATACAAACCAAA 1437
 QY 4474 aacgcccgagtaactgtgggtgatttaaatgcccgttgcacaaacacgcgtgaccttgca 4533
 Db 1438 AACCCGCAAGTAACTGTGGGTGATTTAAATGCCGTTGCCAAACACCCGCTGACCTTTGCA 1497
 QY 4534 ggggatacagggcacaacggttaaaaaactgggcagagctttgacacatcaaaagtgggca 4593
 Db 1498 GGGGATACAGGCACAACGGCTTAAAAAACTGGCGGAGACTTTGACCATCAAAAGGTGGGCAA 1557
 QY 4594 acagacaccaataagctaaacgataataacatcggtgtgtgtagcaggtactgagtggttc 4653
 Db 1558 ACAGACACCAATAGCTTAACCGATATAACATCGGTGTGTGTAGCAGTACTGATGGCTTC 1617
 QY 4654 actgtcaaaacttgccaaagacctaaccaatcttaacagcgttaagtgaagtggaaccaa 4713
 Db 1618 ACTGTCAAACTTGGCAAAAGACTTAACCAATCTTAACAGCGTTAATGAGGTGGGCCCAAA 1677
 QY 4714 attaatgaaaagggcatctcttttgaacgcaaaacggctcaagcacaagcaaacacccct 4773
 Db 1678 AATTGATGACAAAGGCGTGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAACACCCCT 1737
 QY 4774 gtgctaagtgcacaaatggggctggacctgggtgcaaggtcactcaagtaagtgtgggcaagggc 4833
 Db 1738 GTGCTAAGTGCCTAATGGCTGGCTGGCTGGTGCACAAAGTCAATCAGTAATGTGGGCAAGGC 1797
 QY 4834 acaaaagataccgcagctgcgaatgtacaacgttaaacgaagtacgcgaactgttgggt 4893
 Db 1798 ACAAAGATACCCAGCGCTGCCAATGTACACAGTTAAACGAAGTACGCAACTTGTGGGT 1857
 QY 4894 ct-----Lggttaatgataacgctgacgcaatcagggtaaacatttgcgcacatacaa 4944
 Db 1858 CTTGGTAAATGCTGGTAAATGATTAACGCTGACGCAATACAGTAAACATTTGCCGACATCAA 1917
 QY 4945 aaagacccaaatcaggttccatctcaacgcactgtcatcaagcagcagcagcagcagcagc 5004
 Db 1918 AAAGACCCAAATTCAGGTTTCATCTAACCACACTGTCTCAAAAGCAGGACGCTACTT 1977
 QY 5005 ggcggtaaaaggttaataacgataccgaaaaacttgcacactgggttggtgtacaagtgggcg 5064
 Db 1978 GCGGCTAAAGGTAATAACGATACCGAAAAACTTGGCACCTGGTATACAGTGGGCGGTG 2037
 QY 5065 gataaagacggcaacgctcaacggcgttttaagcaatgttgggtcaaaaccccaaaagat 5124
 Db 2038 GATAAAGACGGCAACGCTAACGGCGATTAAAGCAATGTTTGGGTCAAAACCCCAAAAGAT 2097
 QY 5125 ggcagcaaaaaagccctgctgcacactataacgcgcagcaggtcagacccaactatgtgacc 5184
 Db 2098 GGCAGCAAAAAGCCCTGCTCGCCACTTATAAGCGCGCAGGTGAGACCAACTATTGTGACC 2157
 QY 5185 acaaccccgacgaagccattgacagaataaatgaaacaggttatccgcttctccatgac 5244
 Db 2158 AACAAACCCCGCAGAAAGCCATTGACAGATAAATGAACAAGGTATCCGCTTCTTCCATGTC 2217
 QY 5245 aacgatggcaatcaagagcctgtgggtacaagggcgtaaacggcattgactcaagtgacctca 5304
 Db 2218 AACATGGCAATCAAGAGCCCTGTGTACAAAGGCGTAAACGGCATTCACAGTGCCTCA 2277
 QY 5305 ggcaagcactcagtgggcgaatggtttccagggccaaagcagatggtgaagccgcgctggcc 5364
 Db 2278 GGCAGCACTCAGTGGCGATAGGTTTCCAGGCCAAGCAGATGTTGAAGCCCGCGGTGCC 2337

[illegible]

PAKSVWIGFATNDGETNVAIGAKSKSSKAASIAIGDRAKALDNOAIAIGONATANS
WDISIGROAGAEQTEVSAEGRNIAIGDALKRGVNNNIALGTSGDRLAGTHNYLM
GTYVNADEAVRSALTAGTTEKEINAKETDNNKYIEASNTVALGTALATOLAAVIG
QOAKAFNGSVAGNDTASGQOTATGSSNAHATGSSIAIGGTVANGDTPARTLASD
FTIAMGLSAOATKSDAIVAGRNKAAGINTVSYCNAGVTKTADAQSDPNKLEPAT
DAVFIGKAGKSNQRMVSLGDSGEGVGVGTENVITIGSAGKNTKGNNTVAISSRA
QONVEGHDNFAALLEAGONIKGSDNIAIGKHAGRSADPNTKLAINTTISLGKESVSLK
NFGIAQGNKATDGLASIAIGRNAEAVGGTANATIGDSASADAGNIVLGTQAOKS
LTVDGKRYGAYSIVIGTEAKATAAQAAPKNENPKDAIAGTKRAEYASTIALGEG
AKSDTKAQAAGVSGYNSAKGQATAFGESEAKTTENAGSSIAFGTKAOTRASATLGM
CAETGFDGQDGLSDAVALCREAKARONALAFYKAVADHKDAVALGAGATAAE
GTNEATVNEFRYSFAGTKPATVSGVKDAERTITNVAAGRIDKTDIAINSQLYL
ALNALGVNGLTVNLVGGDAIVKEGDEAGTLTMSNIGGTGKTIDHDAEAVNTAK
ASKTIVREGDNLTVBERAAGDSRTYVATKKVDFSVVAGGTFKIDANGLTFFVDOG
TKIDNTPSISIKTGIDAGNKVNTQNGNIANDSKDAVNGQOLFAQGEVKNLIGDPT
YNPETGEYANTIGGTGASTIDEAKAVNTTAKAATEVVOGENIVVYISAPGANGTV
YTVATATVETDKTIVGTVTDKNTDITGLSKNDLGDNPAKNGRAASEEQUNATQ
NLATLGGNAQNTGNVMTDITGKTGNINIDAKASNEVKQKGNMVTPTTGANGQ
TIYEVATADKVADEVKGGTIDATNKISGIAKGDISENSTDAVNGSOLYELQOKI
AKSGDNYNLNRINKVDKDLRAGIAGANAAGLPQAVIPGKSMVAAGAITYKQONAI
ALGWSRISDNGKVIKLTGNTNSRGDFCASIGAGYQW"

BASE COUNT 1320 a 688 c 934 g 958 t
ORIGIN

Query Match 1.5%; Score 92.4; DB 3; Length 3900;
Best Local Similarity 55.8%; Pred. No. 8, 1e-10;
Matches 202; Conservative 0; Mismatches 151; Indels 9; Gaps 1;

QY 5805 tgTggcagcaggtgaggtcagtgccaccagccagcagcgatgcggtcaatggtgtagccagttgta 5864
Db 3534 TATTGCAAAAGGTGACATTTCTGAAACAGTAGTACTGATGCTGTGAATGTTAGCCAGCTATA 3593

QY 5865 caagccaccacaaagcagtgccaaacaaacaaatgagcttgaccat-----cgtat 5915
Db 3594 TGAATTACAAACAAAATCGCTAAATCTGGTGATACTATACATCTCTAAATAACCGTAT 3653

QY 5916 ccaccacaaacgaaataaaagcccaatgcagggatttccatcagcgatggcgatggcgctccat 5975
Db 3654 CAATAAGTGATAAAGATCTCCGTCGGGTATTGACAGCGCGCAATGCACGCGCTGTTT 3713

QY 5976 gccacaagccatcattccctggcagatccatggttaccgggggtattgcccaccacacacg 6035
Db 3714 ACCAAGCGGTATATCCCTGGTAAGAGCATGGTTGCGGTTGTCAGCGCGGCACITACAAAG 3773

QY 6036 tcaaggtgcggtgcagtgaggcagctgcgaagctgcggaataatggtcgaatggtgatttaa 6095
Db 3774 ACNAAAGCAATTGCATTGGTATGTCCTGATTTCTGATAACGGTAAAGTCATTATCAA 3833

QY 6096 aatcaatggttcagccgatacccaagccatgtagggcgcgatggtgtaggagtttca 6155
Db 3834 ATTAACGGGTAAACAAACATCGCGTGTGACTTTGTGTGCAAGCATCGTGCAGGTTATCA 3893

QY 6156 ct 6157
Db 3894 AT 3895

RESULT 14
AE006194/c 10533 bp DNA BCT 08-MAR-2001
LOCUS Pasteurella multocida PM70 section 161 of 204 of the complete genome.
DEFINITION AE006194 AE004439
ACCESSION AE006194.1 GI:12721958
VERSION
KEYWORDS
SOURCE Pasteurella multocida.
ORGANISM Pasteurella multocida.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
REFERENCE 1 (bases 1 to 10533)
AUTHORS May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and Kapur,V.

TITLE Complete genomic sequence of Pasteurella multocida, Pm70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 21145866
REFERENCE 2 (bases 1 to 10533)
AUTHORS Zhang,Q. and Kapur,V.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
55108, USA

FEATURES
source
1..10533 /organism="Pasteurella multocida"
/strain="PM70"
/db_xref="taxon:747"
91..804 /gene="PM1566"
91..804 /gene="PM1566"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03650.1"
/db_xref="GI:12721959"
/translation="MTLKHILWIFLALPIALNAOPVDNKLQHNVOFSAEVVKEIKDK
ELQVTLFLQEEGDASALNQLIIVARNQALDAVKHSTVQVYVSOQRYTOVRVKGEGKO
NGWIDRAEIVLRSQDMPMLSKLVAELNEHFKIGGYATVSEESLAVEREMTEAVLOK
FEQAOLIQSLMKRAGYRLIELDLSPHDVNGSTRSTAVSMKSKDAFYSSADHEMALE
GGTKLRVAVVQAQIELIND"
868..1467 /gene="fkpA"
/note="PM1567"
868..1467 /gene="fkpA"
/codon_start=1
/transl_table=11
/product="fkpA"
/protein_id="AAK03651.1"
/db_xref="GI:12721960"
/translation="MKTAKNVVSYIAVQVRTEDGLVDEAPVNOPLLEYLOGHNNLVIG
LENALRGKAVGDKFEVRVKBEEAYGEYNNVORPKDQFGVDLGVGMRFYADTDI
GPLPVVITEVAENDVVDVGNHMLAGQELLEFSVEVATREATLEIAHGIHQEGCGG
GHHHDSDEEGHCGCGSHHHHEHEHHAHDCCCGGCKH"
complement(1521..1733)
/gene="PM1568"
complement(1521..1733)
/gene="PM1568"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03652.1"
/db_xref="GI:12721961"
/translation="MFPEFDLISKLLKLEDAHFNRLFEKHNLDRHRIKNTAEITELGT
PQEVNKKKELRLKDEIYVILKKHK"
complement(1830..2300)
/gene="yfkJ"
/note="PM1569"
complement(1830..2300)
/gene="yfkJ"
/codon_start=1
/transl_table=11
/protein_id="AAK03653.1"
/db_xref="GI:12721962"
/translation="MTIHLFLVCLGNICRSPMAEFLMREKIKQAGLETVIKTSAGTS
GWHDCGDMHCGTADLDQCHDSTGFRSKVKVQRHQWQFDFIAMDNSNLQDLERLFG
HHPDKLFKITELCPDLGVDPWYSKDFNQTYOLLNPPCCDALLARUREKHL"
complement(2415..6314)
/gene="hsf_2"
complement(2415..6314)
/gene="hsf_2"
/codon_start=1
/transl_table=11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 13:29:49 ; Search time 1369.14 Seconds
(without alignments)
2824.580 Million cell updates/sec

Title: US-09-361-619-8
Perfect score: 6159
Sequence: 1 atgaatcacatcataaagt.....ttggtgcagggttttcacttt 6159

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6159	100.0	6159	22	AAF59104
2	4101.6	66.6	6144	22	AAF59103
3	4101.6	66.6	6972	22	AAF59102
4	4090.6	66.4	6971	17	AAT38740
5	4090.6	66.4	6973	22	AAF59100
6	3982	64.7	62909	22	AAF28545
7	3947.8	64.1	5979	22	AAF59101
8	3946.8	64.1	6259	22	AAF59106
9	2939.2	47.7	6942	22	AAF59105
10	166.2	2.7	720	22	AAF59129
11	144.6	2.3	3030	21	AA92496
					M. catarrhalis str
					M. catarrhalis str
					M. catarrhalis str
					Moraxella outer me
					M. catarrhalis str
					Genomic fragment #
					M. catarrhalis M56
					M. catarrhalis str
					M. catarrhalis str
					M. catarrhalis str
					Haemophilus influe

12	143	2.3	3036	21	AA92493	Haemophilus influe
13	138.2	2.2	3354	21	AA92495	Haemophilus influe
14	129.4	2.1	3342	21	AA92498	Haemophilus influe
C 15	122	2.0	936	22	AAF58252	Oligonucleotide D1
C 16	122	2.0	936	22	AAF58254	Oligonucleotide D1
C 17	122	2.0	936	22	AAF58257	Oligonucleotide D1
C 18	122	2.0	936	22	AAF58259	Oligonucleotide D2
C 19	122	2.0	936	22	AAF58262	Oligonucleotide D2
C 20	122	2.0	936	22	AAF58255	Oligonucleotide D1
21	119.2	1.9	936	22	AAF58252	Oligonucleotide D1
22	119.2	1.9	936	22	AAF58254	Oligonucleotide D1
23	119.2	1.9	936	22	AAF58257	Oligonucleotide D1
24	119.2	1.9	936	22	AAF58259	Oligonucleotide D2
25	119.2	1.9	936	22	AAF58262	Oligonucleotide D2
26	119.2	1.9	938	22	AAF58255	Oligonucleotide D1
27	84.4	1.4	3294	17	AAT41475	Haemophilus adhesi
28	83.6	1.4	2037	17	AAT41477	Haemophilus adhesi
29	82	1.3	2079	21	AA92494	Haemophilus influe
30	81.8	1.3	7486	19	AAV22837	Haemophilus paraga
31	80.2	1.3	8930	19	AAV22834	Haemophilus influe
32	77.2	1.3	7253	21	AA92499	Haemophilus influe
33	77.2	1.3	7291	17	AAT41476	Haemophilus adhesi
34	76.4	1.2	3300	21	AA92497	Haemophilus influe
C 35	75	1.2	1830121	17	AA742063	Haemophilus influe
36	65.2	1.1	1757	21	AA92500	Haemophilus influe
37	60	1.0	60	17	AAT38741	Moraxella outer me
38	58.4	0.9	1797	20	AA85794	DNA encoding a sur
39	56.8	0.9	1797	20	AA85790	DNA encoding a sur
40	55.8	0.9	1779	20	AA85790	DNA encoding N. me
41	55.4	0.9	3567	21	AA70117	Plasmodium falcipa
42	55.2	0.9	1770	20	AA85797	DNA encoding a sur
43	55.2	0.9	1776	20	AA85798	DNA encoding N. me
44	55.2	0.9	1776	20	AA85798	DNA encoding a sur
45	55.2	0.9	1776	20	AA85793	DNA encoding a sur

ALIGNMENTS

RESULT 1

AAF59104
ID AAF59104 standard; DNA; 6159 BP.

XX AAF59104;

AC AAF59104;

XX 24-APR-2001 (first entry)

XX M. catarrhalis strain Q8 200kDa gene SEQ ID NO:8.

DE Moraxella catarrhalis

XX Moraxella catarrhalis strain Q8; major outer membrane protein;
200kDa outer membrane protein; antibacterial; immunogenic; infection;
otitis media; detection; ds.

XX Moraxella catarrhalis.

OS Moraxella catarrhalis.

XX WO200107619-Al.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX P-PSDB; AAB69135.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
useful in protective vaccines and for diagnosis

XX

Claim 1; Fig 4A-V; 247pp; English.

PS The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (Ab), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis strain Q8 200kDa gene, which is given in the CC exemplification of the present invention.

XX Sequence 6159 BP; 2035 A; 1386 C; 1385 G; 1353 T; 0 other;

Query Match 100.0%; Score 6159; DB 22; Length 6159; Best Local Similarity 100.0%; Pred. No. 0; Matches 6159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgaatcacatctataaagtoatctttaacaaagccacagccacatttatgccgtggcg	60
Db	1	atgaatcacatctataaagtoatctttaacaaagccacagccacatttatgccgtggcg	60
Qy	61	gaatatccaaatccacagtcacgggggggtagctgtgtctacagggcaagtggcag	120
Db	61	gaatgcccacatccacagtcacgggggggtagctgtgtctacagggcaagtggcag	120
Qy	121	gtacgcacttaagtgttgcgcgtattgcgcgtctgcgtctctcgtgcgtgcgcgc	180
Db	121	gtacgcacttaagtgttgcgcgtattgcgcgtctgcgtctctcgtgcgtgcgcgc	180
Qy	181	ctcaatggcagtcgtattgtctcaacaaattactaccaagatcgaaattgttcaacaaac	240
Db	181	ctcaatggcagtcgtattgtctcaacaaattactaccaagatcgaaattgttcaacaaac	240
Qy	241	aagataaacacacgctgaaggcgatccctagcgacagtgagcatcattgctttt	300
Db	241	aagataaacacacgctgaaggcgatccctagcgacagtgagcatcattgctttt	300
Qy	301	ggtagctttctaaagcacaagctctcaagctattgtctatcgttagtgcacaccagat	360
Db	301	ggtagctttctaaagcacaagctctcaagctattgtctatcgttagtgcacaccagat	360
Qy	361	cctaataatggtagtaattgtagttaggttccacgcacaaaggtaacgagtcctgcgc	420
Db	361	cctaataatggtagtaattgtagttaggttccacgcacaaaggtaacgagtcctgcgc	420
Qy	421	atcggttgtagtattgctgaggtgtagtgcctgattgctcgttagtgcagctta	480
Db	421	atcggttgtagtattgctgaggtgtagtgcctgattgctcgttagtgcagctta	480
Qy	481	tatttgcttaagaatcttgaatctgaagaatgaattccacaaacttattcattgcccata	540
Db	481	tatttgcttaagaatcttgaatctgaagaatgaattccacaaacttattcattgcccata	540
Qy	541	atatataaaaaatacaaacctcaacgattgtaaaatcaaatatcagcagcaagagca	600
Db	541	atatataaaaaatacaaacctcaacgattgtaaaatcaaatatcagcagcaagagca	600
Qy	601	caagggcacgcagtagtgcagtgaggccatgtcatatgcacaggggtcattttccaac	660
Db	601	caagggcacgcagtagtgcagtgaggccatgtcatatgcacaggggtcattttccaac	660
Qy	661	gcctttggtacacgacagctgaagctgcctcttccttggcagtaggtcttgcgcgc	720
Db	661	gcctttggtacacgacagctgaagctgcctcttccttggcagtaggtcttgcgcgc	720

Qy	721	caagccacaaaaaatcttcaatcgctgttggttccaatgcaaaagctaaacgcgtttgca	780
Db	721	caagccacaaaaaatcttcaatcgctgttggttccaatgcaaaagctaaacgcgtttgca	780
Qy	781	ggacagccatttggggaaatacttagttaattgggtcgagcggttgccttaggtttt	840
Db	781	ggacagccatttggggaaatacttagttaattgggtcgagcggttgccttaggtttt	840
Qy	841	ggttccagatccttgatagggataataatagatgcagtcgagtcctatgtaccactaggt	900
Db	841	ggttccagatccttgatagggataataatagatgcagtcgagtcctatgtaccactaggt	900
Qy	901	aaaacgttagcagcagcagataaaagccaccgccaggggtgattctcagatattttcc	960
Db	901	aaaacgttagcagcagcagataaaagccaccgccaggggtgattctcagatattttcc	960
Qy	961	attggttaataatagtaataatagcagtcagtcagcgctaaatcatcatcgtcgtg	1020
Db	961	attggttaataatagtaataatagcagtcagtcagcgctaaatcatcatcgtcgtg	1020
Qy	1021	ggttctcgggataccgagtcggtcgaattgtggcacagcttaattggtggaggaactg	1080
Db	1021	ggttctcgggataccgagtcggtcgaattgtggcacagcttaattggtggaggaactg	1080
Qy	1081	aatcgtaaaattacttttaaggtgtagtgacaataatagcaatagcgttagaagaaggt	1140
Db	1081	aatcgtaaaattacttttaaggtgtagtgacaataatagcaatagcgttagaagaaggt	1140
Qy	1141	ttgggcaataactttaactattaaaggtgtagtcacagacccaagcatttaaccgaagctaac	1200
Db	1141	ttgggcaataactttaactattaaaggtgtagtcacagacccaagcatttaaccgaagctaac	1200
Qy	1201	atcggtggtgaacagatgcaatggtctctgaaagttaaaactgtcaagagctgactgga	1260
Db	1201	atcggtggtgaacagatgcaatggtctctgaaagttaaaactgtcaagagctgactgga	1260
Qy	1261	ttgacagctgtctcgcgtaccacaaataatcccggttagtaataaccaacaacaacgccc	1320
Db	1261	ttgacagctgtctcgcgtaccacaaataatcccggttagtaataaccaacaacaacgccc	1320
Qy	1321	gagctacaagaagcgtggtttgaccttttagcccaataacaggttacaacacagataaaacc	1380
Db	1321	gagctacaagaagcgtggtttgaccttttagcccaataacaggttacaacacagataaaacc	1380
Qy	1381	gtctacagcatttgatggttagttagttagttagttagttagttagttagttagttagttagt	1440
Db	1381	gtctacagcatttgatggttagttagttagttagttagttagttagttagttagttagttagt	1440
Qy	1441	actactcgtattaccacaaagaataatggtttgctggttactaatgaggttgatgaa	1500
Db	1441	actactcgtattaccacaaagaataatggtttgctggttactaatgaggttgatgaa	1500
Qy	1501	agcaaaccttatttgacacaaagcgttaaaagtgtggcaacagcagccctaaacagctggt	1560
Db	1501	agcaaaccttatttgacacaaagcgttaaaagtgtggcaacagcagccctaaacagctggt	1560
Qy	1561	agcttgactgttaataacacacactgtaataacaaatccaagtcggtgcttaattggcatt	1620
Db	1561	agcttgactgttaataacacacactgtaataacaaatccaagtcggtgcttaattggcatt	1620
Qy	1621	aaatttgcacagtcgctaaataatgttgcaaatctctcagcaacagctcggcactgctct	1680
Db	1621	aaatttgcacagtcgctaaataatgttgcaaatctctcagcaacagctcggcactgctct	1680
Qy	1681	attaccgaagaagaataatggtttgctggttactaatgaggttgatgaacagcagca	1740
Db	1681	attaccgaagaagaataatggtttgctggttactaatgaggttgatgaacagcagca	1740
Qy	1741	tatttgataaagaacgacttaaaagtgtggttggatgaaattaccacagatagtggtatt	1800
Db	1741	tatttgataaagaacgacttaaaagtgtggttggatgaaattaccacagatagtggtatt	1800
Qy	1801	aatgctggttaatacaagaattaccggacttacttaattggttagcaaataccgaatgcggtt	1860

QY 4021 ggcgtataaacaccacattgaccagtactgtgcagagtgcataataattggcctaagc 4080
DB 4021 ggcgtataaacaccacattgaccagtactgtgcagagtgcataataattggcctaagc 4080
QY 4081 aatcaagctactggcgatgcgtttgtcaaggccagtgatatacgttgctcatctaaacacc 4140
DB 4081 aatcaagctactggcgatgcgtttgtcaaggccagtgatatacgttgctcatctaaacacc 4140
QY 4141 ttatctggcgacatccaactgccaaaggcgcaagccgaaggaacaaactcagcggtcat 4200
DB 4141 ttatctggcgacatccaactgccaaaggcgcaagccgaaggaacaaactcagcggtcat 4200
QY 4201 gtgagtgcgtggcgaataaggtctatctgtacagtaccgataaacaagtactataacacc 4260
DB 4201 gtgagtgcgtggcgaataaggtctatctgtacagtaccgataaacaagtactataacacc 4260
QY 4261 aaaaaatgatggcagttgataaaaacaaagaagtggccaaagacaaaactggcgcccaa 4320
DB 4261 aaaaaatgatggcagttgataaaaacaaagaagtggccaaagacaaaactggcgcccaa 4320
QY 4321 gcccaaacccagatggcatttgggtccaaatgaattgtcaaatcagtcatttaacaaagaa 4380
DB 4321 gcccaaacccagatggcatttgggtccaaatgaattgtcaaatcagtcatttaacaaagaa 4380
QY 4381 caagtataatgatcccaataaaaaagcagcatcaatgaagacaaaccccttggtaaaagga 4440
DB 4381 caagtataatgatcccaataaaaaagcagcatcaatgaagacaaaccccttggtaaaagga 4440
QY 4441 cttgaaaagccgtctctgataaaaacaaacccaaacgcgcagtaactgtgggtgattta 4500
DB 4441 cttgaaaagccgtctctgataaaaacaaacccaaacgcgcagtaactgtgggtgattta 4500
QY 4501 aatgccgttggccaaacacccgtgcacotttgcaggggatcacaggcacacgcgtataaaaa 4560
DB 4501 aatgccgttggccaaacacccgtgcacotttgcaggggatcacaggcacacgcgtataaaaa 4560
QY 4561 ctggcgagactttgaccatacaagggtgggcaaacacagacaccataaagctaaacgataat 4620
DB 4561 ctggcgagactttgaccatacaagggtgggcaaacacagacaccataaagctaaacgataat 4620
QY 4621 aacatcgggttggtagcaggtactgtagtgccttcaacttcaaaacttgcgaagacctaaacc 4680
DB 4621 aacatcgggttggtagcaggtactgtagtgccttcaacttcaaaacttgcgaagacctaaacc 4680
QY 4681 aatcttaacagcgttaatgcaggtggcaccaaaattgatgaaaaagcattctcttttga 4740
DB 4681 aatcttaacagcgttaatgcaggtggcaccaaaattgatgaaaaagcattctcttttga 4740
QY 4741 gacgcacacgggtcaagccaaagcaaacccctgtgctaagtgcataatggcgtagacctg 4800
DB 4741 gacgcacacgggtcaagccaaagcaaacccctgtgctaagtgcataatggcgtagacctg 4800
QY 4801 ggtgcaaggtcatcagtaattgtggcaaaaggcacaagaatgacgcagctgcgaatgta 4860
DB 4801 ggtgcaaggtcatcagtaattgtggcaaaaggcacaagaatgacgcagctgcgaatgta 4860
QY 4861 caacagttaaacgaagtagcgaacttgggtcttgggtatgataacagctgcagcggaat 4920
DB 4861 caacagttaaacgaagtagcgaacttgggtcttgggtatgataacagctgcagcggaat 4920
QY 4921 caggtaaacattgcccacatcaaaaagacccaaatcagggttcacatctaaacgcgact 4980
DB 4921 caggtaaacattgcccacatcaaaaagacccaaatcagggttcacatctaaacgcgact 4980
QY 4981 gtcatcaaacgagcgacttacttggcggttaagggtataacagataccgaaaaacttgc 5040
DB 4981 gtcatcaaacgagcgacttacttggcggttaagggtataacagataccgaaaaacttgc 5040
QY 5041 actcgttggtgtacaagtggcggtgataaagcagcgcaacgcgttaacggcgatttaagcaat 5100
DB 5041 actcgttggtgtacaagtggcggtgataaagcagcgcaacgcgttaacggcgatttaagcaat 5100

QY 5101 gtttgggtcaaaaccccaaaaagatggcagcaaaaaagccctgtcgcaccttataacgcc 5160
DB 5101 gtttgggtcaaaaccccaaaaagatggcagcaaaaaagccctgtcgcaccttataacgcc 5160
QY 5161 gcaggttcagacccaactatgtgaccacaaccccgcagagccattgacagataaataagaa 5220
DB 5161 gcaggttcagacccaactatgtgaccacaaccccgcagagccattgacagataaataagaa 5220
QY 5221 caaggtatccgcttctccatgtcaacgatggcaatcaagagccctgtgtgtaacaaaggcct 5280
DB 5221 caaggtatccgcttctccatgtcaacgatggcaatcaagagccctgtgtgtaacaaaggcct 5280
QY 5281 aacggcattgactcaagtgccctcagcgaacgactcagtgaggcataggtttccagggccaag 5340
DB 5281 aacggcattgactcaagtgccctcagcgaacgactcagtgaggcataggtttccagggccaag 5340
QY 5341 gcagatgtgaagccgcgttgcataaggcagacaaccccaagcaggaaccaatccatc 5400
DB 5341 gcagatgtgaagccgcgttgcataaggcagacaaccccaagcaggaaccaatccatc 5400
QY 5401 gccatcgtgtataacgcacaaagccacggggcgatcaatccatgccatcggtacagggcaat 5460
DB 5401 gccatcgtgtataacgcacaaagccacggggcgatcaatccatgccatcggtacagggcaat 5460
QY 5461 gtggtagcaggtaaagcactctgttgcctcggcgaccccaagcactgttaaggctgataac 5520
DB 5461 gtggtagcaggtaaagcactctgttgcctcggcgaccccaagcactgttaaggctgataac 5520
QY 5521 agttacagtggtgtataacaaacccagttaccgatgccactcaaacccgatgtcttgggt 5580
DB 5521 agttacagtggtgtataacaaacccagttaccgatgccactcaaacccgatgtcttgggt 5580
QY 5581 gtgggcaataacatacaccgtgcacgaagaatgaactcgttgccttaggttcaaaactcgcc 5640
DB 5581 gtgggcaataacatacaccgtgcacgaagaatgaactcgttgccttaggttcaaaactcgcc 5640
QY 5641 atcagtgaggcacacacgcagggcacacaagcccaaaaatctgcagcgacagcggtatac 5700
DB 5641 atcagtgaggcacacacgcagggcacacaagcccaaaaatctgcagcgacagcggtatac 5700
QY 5701 accacacagcaggtgacacaggtacggttaaggcttctgtggcaaacacgcggttgggt 5760
DB 5701 accacacagcaggtgacacaggtacggtttaaaggcttctgtggcaaacacgcggttgggt 5760
QY 5761 gcggtctccgtgggtgcctcaggtgctgaacgccgtatccaaaaatgtggcgagcggtgag 5820
DB 5761 gcggtctccgtgggtgcctcaggtgctgaacgccgtatccaaaaatgtggcgagcggtgag 5820
QY 5821 gtdagtgccacacagcagcagatgcggtcaatgtagccagttgtacaagccaccacaagc 5880
DB 5821 gtdagtgccacacagcagcagatgcggtcaatgtagccagttgtacaagccaccacaagc 5880
QY 5881 attgccaacgcgaaccaatgagcttgaccatcgatccacaaaacgaaaaataaagccaat 5940
DB 5881 attgccaacgcgaaccaatgagcttgaccatcgatccacaaaacgaaaaataaagccaat 5940
QY 5941 gcagggatttcatcagcgatggcgatggcgatggcgatggcgatggcgatggcgatggcgat 6000
DB 5941 gcagggatttcatcagcgatggcgatggcgatggcgatggcgatggcgatggcgatggcgat 6000
QY 6001 tccatggttaccgggggtattgccaccccaacacggttcaagggtgcggtggcgatggcgat 6060
DB 6001 tccatggttaccgggggtattgccaccccaacacggttcaagggtgcggtggcgatggcgat 6060
QY 6061 tcgaagctgcggataatggtcaatgggtatttaaaatcaatgggttcagcgcgatacccaa 6120
DB 6061 tcgaagctgcggataatggtcaatgggtatttaaaatcaatgggttcagcgcgatacccaa 6120
QY 6121 ggcacatgagggcgaggtgtgaggggttttcacttt 6159
DB 6121 ggcacatgagggcgaggtgtgaggggttttcacttt 6159

Db 3394 atcagcgtactgcgaacacgacaaactcattcacgaattctcagtagcagatgaa 3453
Qy 3481 caagtgtaactttagctggttagtaaaccttactccagttatgacacctcaaaagacctct 3540
Db 3454 caagtgtaactttagctggttagtaaaccttactccagttatgacacctcaaaagacctct 3513
Qy 3541 gatgcatcaacttgcaggtgaaacggaatcaccaccaggttaataaagggtggtggtg 3600
Db 3514 gatgcatcaacttgcaggtgaaacggaatcaccaccaggttaataaagggtggtggtg 3573
Qy 3601 cgtggtgacttagcaacacaaaggcttaaccacgcttaagctgacgtgggttaataat 3660
Db 3574 cgtggtgacttagcaacacaaaggcttaaccacgcttaagctgacgtgggttaataat 3633
Qy 3661 aatggcaaggacttgcatttaacagccaaatggtcaaaaataccatcacaggactaagc 3720
Db 3634 aatggcaaggacttgcatttagacagccaaatggtcaaaaataccatcacaggactaagc 3693
Qy 3721 aacactctagctaatgttaccatgataaaggtagcgtacgaccacagaaagggcaat 3780
Db 3694 aacactctagctaatgttaccatgataaaggtagcgtacgaccacagaaagggcaat 3753
Qy 3781 ataataaagacagaacaaacccgctgcgcgaagcttgttgatgtgctaaagcgcaggc 3840
Db 3754 ataataaagacagaacaaacccgctgcgcgaagcttgttgatgtgctaaagcgcaggc 3813
Qy 3841 tttaacttgaaggcaatggtgaagcgttgacttcttccacttatgacacgcgtcaac 3900
Db 3814 tttaacttgaaggcaatggtgaagcgttgacttcttccacttatgacacgcgtcaac 3873
Qy 3901 ttgcgaatggcaataccacacccgctaaggtagaccttatgatgacacagcaaaacccagt 3960
Db 3874 ttgcgaatggcaatggcaacacccgctaaggtagaccttatgatgacacagcaaaacccagt 3933
Qy 3961 aaagtgtctatgatgctcaatgtgagatgatacaacacattaaagttaagataaaaaactt 4020
Db 3934 aaagtgtctatgatgctcaatgtgagatgatacaacacattaaagttaagataaaaaactt 3993
Qy 4021 ggcgtaaaaacccacacattgaccagctactggcacaggtgctataaatttgcacctaaagc 4080
Db 3994 ggcgtaaaaacccacacattgaccagctactggcacaggtgctataaatttgcacctaaagc 4053
Qy 4081 aatacagctactgcgcatgagcttgcgaagccagtgatctcgttgcctatctataaacacc 4140
Db 4054 aatacagctactgcgcatgagcttgcgaagccagtgatctcgttgcctatctataaacacc 4113
Qy 4141 ttatctggcagatcccaactgccaagggcgaagccaaagcgaacgaactcagcaggctat 4200
Db 4114 ttatctggcagatcccaactgccaagggcgaagccaaagcgaacgaactcagcaggctat 4173
Qy 4201 gtgcatgctgagtgcaataaagtgctatctatgacagttaccgtatacaagctataagcc 4260
Db 4174 gtgcatgctgagtgcaataaagtgctatctatgacagttaccgtatacaagctataagcc 4233
Qy 4261 aaaaatgatgcacagtgatgataaaaccaaagaagtgccaaaagcaaaactggtgcgcca 4320
Db 4234 aaaaatgatgcacagtgatgataaaaccaaagaagtgccaaaagcaaaactggtgcgcca 4293
Qy 4321 gcccaaacccagatggcacattggctcaaatgaatgctcaaatcagtcatttaacaaagaa 4380
Db 4294 gcccaaacccagatggcacattggctcaaatgaatgctcaaatcagtcatttaacaaagaa 4353
Qy 4381 caagtaaatgatgcataaaagaagcgaatcaatgaagaacacgccttgggttaaaagga 4440
Db 4354 caagtaaatgatgcataaaagaagcgaatcaatgaagaacacgccttgggttaaaagga 4413
Qy 4441 cttgaaaagccgcttctgataacaaacccaaagccgcagtaacttgggttgattta 4500
Db 4414 cttgaaaagccgcttctgataacaaacccaaagccgcagtaacttgggttgattta 4473
Qy 4501 aatccggttgcgaacacacgcgtgaccttgcaggggatacaggcacaacggcttaaaaaa 4560
Db 4501 aatccggttgcgaacacacgcgtgaccttgcaggggatacaggcacaacggcttaaaaaa 4560

Db 4474 aatccggttgcgaacacacgcgtgaccttgcaggggatacaggcacaacggcttaaaaaa 4533
Qy 4561 ctggcgcagactttagaccatcaaaaggtggcaaacagacacaaataagcttaacgataat 4620
Db 4534 ctggcgcagactttagaccatcaaaaggtggcaaacagacacaaataagcttaacgataat 4593
Qy 4621 aacatcgggtggtgtagcaggttactgagttcactgtcgaacttgcgaagaccataacc 4680
Db 4594 aacatcgggtggtgtagcaggttactgagttcactgtcgaacttgcgaagaccataacc 4653
Qy 4681 aatcttaacagcgttaattgacaggtggcaccaaaattgtgaaaaagcgtctctttgata 4740
Db 4654 aatcttaacagcgttaattgacaggtggcaccaaaattgtgaaaaagcgtctctttgata 4713
Qy 4741 gacgcaaacggtcgaagccaaagcaaacacccctgtgctaaagtgcgaatgggctgacctg 4800
Db 4714 gactcaagcgtcgaagccaaagcaaacacccctgtgctaaagtgcgaatgggctgacctg 4773
Qy 4801 ggtggcaaggtcactagtaattgtgggcaaaaggcacaagaatataccgcgtgcgaatga 4860
Db 4774 ggtggcaaggtcactagtaattgtgggcaaaaggcacaagaatataccgcgtgcgaatga 4833
Qy 4861 caacagttaaacgaagtacgcaacttgttggtctt-----tggttaagtataacgct 4911
Db 4834 caacagttaaacgaagtacgcaacttgttggtcttggtaagtctggttaagtataacgct 4893
Qy 4912 gacggcaatcaggttaaacatttgcgcacatcaaaaagaccccaaaattcaggttcatcatct 4971
Db 4894 gacggcaatcaggttaaacatttgcgcacatcaaaaagaccccaaaattcaggttcatcatct 4953
Qy 4972 aacgcgactgtcactaaagcagcgacgttacttggcggttaaaggtaataacagataccgaa 5031
Db 4954 aacgcgactgtcactaaagcagcgacgttacttggcggttaaaggtaataacagataccgaa 5013
Qy 5032 aaacttgcactgtggtgtacaaagtggcgtgataaagcgaacgcgaacgaacgcat 5091
Db 5014 aaacttgcactgtggtgtacaaagtggcgtgataaagcgaacgcgaacgcat 5073
Qy 5092 ttaagcaatgttgggtcacaaccccaaaagatggcagcaaaaagcctgtcgcacct 5151
Db 5074 ttaagcaatgttgggtcacaaccccaaaagatggcagcaaaaagcctgtcgcacct 5133
Qy 5152 tataacgcgcaggtcagacccaactatgtgaccaacaaccccgacagaagcgaatgacaga 5211
Db 5134 tataacgcgcaggtcagacccaactatgtgaccaacaaccccgacagaagcgaatgacaga 5193
Qy 5212 ataatgaacaggtatccgtcttccatgtcaacagtgccaatcaagagcgtgtgata 5271
Db 5194 ataatgaacaggtatccgtcttccatgtcaacagtgccaatcaagagcgtgtgata 5253
Qy 5272 caagggcgtaacggtgacttgaactcgaagtcctcaggaagcactcagtgatagtttc 5331
Db 5254 caagggcgtaacggtgacttgaactcgaagtcctcaggaagcactcagtgatagtttc 5313
Qy 5332 caggccaagggcagatggtgaagcgccttgcctatagcagacaaccccaagcaggaac 5391
Db 5314 caggccaagggcagatggtgaagcgccttgcctatagcagacaaccccaagcaggaac 5373
Qy 5392 caatccatcgcctatcgtgataaagcgaacccagcgcgtatcaatccatcgcctatcgt 5451
Db 5374 caatccatcgcctatcgtgataaagcgaacccagcgcgtatcaatccatcgcctatcgt 5433
Qy 5452 acaggcaatgtggttagcaggttaagcactctgtgctcctcggcgaacccagcgttgaag 5511
Db 5434 acaggcaatgtggttagcaggttaagcactctgtgctcctcggcgaacccagcgttgaag 5493
Qy 5512 gctgataacagttacagtggtgggttaataacacccagtttaacagtgccactcaaacagat 5571
Db 5494 gctgataacagttacagtggtgggttaataacacccagtttaacagtgccactcaaacagat 5553
Qy 5572 gctcttgggtgggcaataaataccatccgtagcgaagtaactcgttgccttagttca 5631
Db 5554 gctcttgggtgggcaataaataccatccgtagcgaagtaactcgttgccttagttca 5613

[illegible][illegible]

[illegible]

Db 5555 aaacttgccactggtggtacacagtggtggtataaagacggaacgcgtacacggcat 5614
QY 5092 ttaagcaatgttgggtcacaaccccaaaagatggcagcaaaaagccctgctcgcacat 5151
Db 5615 ttaagcaatgttgggtcacaaccccaaaagatggcagcaaaaagccctgctcgcacat 5674
QY 5152 tataacgcccaggtcagacccaactatgtgaccaacaacccccgcagaagccattgacaga 5211
Db 5675 tataacgcccaggtcagacccaactatttgaccaacaacccccgcagaagccattgacaga 5734
QY 5212 ataaatgaacaaggtaccgcttctccatgtcaacgatggcaatcaaaagccctggtga 5271
Db 5735 ataaatgaacaaggtaccgcttctccatgtcaacgatggcaatcaaaagccctggtga 5794
QY 5272 caagggcgtaacgctactcaagtgcctcagggaagcactcagtggcgataggttcc 5331
Db 5795 caagggcgtaacgctactcaagtgcctcagggaagcactcagtggcgataggttcc 5854
QY 5332 caggccaaagcagatggtgaagccgcgttgccatagggcagacaacccccagcaggcaac 5391
Db 5855 caggccaaagcagatggtgaagccgcgttgccatagggcagacaacccccagcaggcaac 5914
QY 5392 caatccatgccatcggtgataacgacacagccacggcgcatcaatccatcgccatcggt 5451
Db 5915 caatccatgccatcggtgataacgacacagccacggcgcatcaatccatcgccatcggt 5974
QY 5452 acagggcaatgtgtagcagtagcactctggtgccatcgccgacccaagcactgttaag 5511
Db 5975 acagggcaatgtgtagcagtagcactctggtgccatcgccgacccaagcactgttaag 6034
QY 5512 gctgataacagttacagtggtggtgataataacacacagtttaccgatgccactcaaacgat 5571
Db 6035 gctgataacagttacagtggtggtgataataacacacagtttaccgatgccactcaaacgat 6094
QY 5572 gcttggtgtggcaataacatcacgtgaacggaagtaactcgttgcttaggttca 5631
Db 6095 gcttggtgtggcaataacatcacgtgaacggaagtaactcgttgcttaggttca 6154
QY 5632 aactctgccatcagtcaggcagcacacgagcagggcacacagcccaaaaatctgcaggcaca 5691
Db 6155 aactctgccatcagtcaggcagcacacgagcagggcacacagcccaaaaatctgcaggcaca 6214
QY 5692 gcagggtacacaccacacagcaggtgccacaggttacggttaaaagcttctgtgacaacgc 5751
Db 6215 gcagggtacacaccacacagcaggtgccacaggttacggttaaaagcttctgtgacaacgc 6274
QY 5752 gcgggttggtgcgtctccgtggtgctcaggtgctgaacgcccgtatccaaaatgtggca 5811
Db 6275 gcgggttggtgcgtctccgtggtgctcaggtgctgaacgcccgtatccaaaatgtggca 6334
QY 5812 gcagggtgaggtcagtcagccacgacacagtcaggtgctcaatggtgacccagttgtacaaagcc 5871
Db 6335 gcagggtgaggtcagtcagccacgacacagtcaggtgctcaatggtgacccagttgtacaaagcc 6394
QY 5872 acccaagaattgccaaagcaacaaatgaactgacatcgatccaccacaaaacgaaaat 5931
Db 6395 acccaagaattgccaaagcaacaaatgaactgacatcgatccaccacaaaacgaaaat 6454
QY 5932 aaagccaatcagggatttcatcagcagatggcagatggcgatccatgccacagccatcatt 5991
Db 6455 aaagccaatcagggatttcatcagcagatggcagatggcgatccatgccacagccatcatt 6514
QY 5992 cctggcagatccatggttaccgggggtattggcaccacacacagcgtgaagtgcggtggca 6051
Db 6515 cctggcagatccatggttaccgggggtattggcaccacacacagcgtgaagtgcggtggca 6574
QY 6052 gtgggactgtcgaagctgtcgggataatggtcaatggttataaatcaatggttccagcc 6111
Db 6575 gtgggactgtcgaagctgtcgggataatggtcaatggttataaatcaatggttccagcc 6634
QY 6112 gatacccaagggccatgtagggcgccaggttggtgcaggttttcacttt 6159
|||||

Db 6635 gatacccaagcccatgtagggcgccaggttggtgcaaggttttcacttt 6682
RESULT 4
AAT38740
ID AAT38740 standard; DNA; 6971 BP.
XX
AC AAT38740;
XX
DT 25-JAN-1997 (first entry)
XX
DE Moraxella outer membrane protein gene.
XX
KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;
diagnosis; ss.
XX
OS Mycobacterium catarrhalis strain 4223.
XX
FH Key Location/Qualifiers
CDS 706..684
FT /*tag= a
XX
PN WO9634960-A1.
XX
PD 07-NOV-1996.
XX
PF 29-APR-1996; 96WO-CA00264.
XX
PR 26-MAR-1996; 96US-0621944.
PR 01-MAY-1995; 95US-0431718.
PR 07-JUN-1995; 95US-0478370.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX
DR WPI; 1996-506162/50.
DR P-PSDB; AAW04505.
XX
PT Moraxella outer membrane protein - useful as immunogen in protective
vaccine and for diagnosis
XX
PS Claim 14; Fig 6; 109pp; English.
XX
CC A genomic DNA fragment (T38740) of Moraxella catarrhalis otitis
media strain 4223 includes the coding region for a 1992-amino
acid protein (W04505) identified as an approx. 200 kDa outer
membrane protein (OMP). The DNA was isolated from a strain 4223
genomic library in phage lambda EMBL3 by screening with an
anti-200 kDa protein guinea pig antiserum. The gene can be
used for the recombinant expression of the OMP (for use in
vaccines), for the prepn. of hybridisation probes, or may be
incorporated into a live vector for use in direct immunisation.
XX
SQ Sequence 6971 BP; 2265 A; 1553 C; 1533 G; 1620 T; 0 other;

Query Match 66.4%; Score 4090.6; DB 17; Length 6971;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;

QY 1 atgaatcacatctataaagtcatttttaacaaagccacagggcacatttatggcgtggcg 60
|||||
Db 540 atgaatcacatctataaagtcatttttaacaaagccacagggcacatttatggcgtggcg 599
QY 61 gaatatcccaaatccccacagtcac-ggggggggtagctgtctacagggcaagttggcag 119
|||||
Db 600 gagtacgcaaatccccacagtcac-ggggggggtagctgtctacagggcaagttggcag 659
QY 120 tgtacgcaacttaagctttgccgtatttgcgcgctcgtctcctgctgacgtggcag 179
|||||
Db 660 tgtatgcactctgagctttgcccgtatttgcgcgctcgtctcctgctgacgtggcag 719

Db	1767	agttaaacttgctaaaaactttaacaatcttactgagtggaatacaactacattaaatgc	1826
Qy	1278	taccaacaaatcacccgttagtaataaccaacaacaacaaacgccgagctcaaaagcggtg	1337
Db	1827	cacaacacagttgaagtaggtagtagtagtagtagtagtagtagtagtagtagtagtagtag	1886
Qy	1338	tttgaccttttagccca-----atacagggtacaaaaacagataaaaacccgttccagcat	1391
Db	1887	tttaacctttaccagcccaatcacggcagtcacaagcagaacagcaaaacccgtctatggcgt	1946
Qy	1392	tgatgattgaaatttactaatgatatgtagtagcaactaaaggtagtactactcgat	1451
Db	1947	taatgggggtgaatttactaataatgcagaacaacagcagcaatcgcgactactcgat	2006
Qy	1452	taccaaaagaaaaattgggtttgctggttactaatgatggagttgatgaagcaaacctta	1511
Db	2007	taccagagataaaattggctttgctcgagatggtagttgatgaaaaaagaagcaccata	2066
Qy	1512	tcttgacaacgaaaaagtgaagtgtggaacagcaccttaaacagtggttagcttgactgt	1571
Db	2067	tttgataaaaaaacacttaagtgggttagttgcaattaccatagacaaatggcattga	2126
Qy	1572	taataaacaccactggttaataaaacaaatccaagtcggtgctaatggcattaaattggcac	1631
Db	2127	tgaggtaataaaaagatcagtaactctgccaaggtagcagtgccaagatgcgggttac	2186
Qy	1632	agtcgctataaatgtgcataactctcagcaacagtcggcactgctcgtattaccogaaga	1691
Db	2187	catcgaaacagctcaaaagccgcgaagctactttaaacgcaggcgctggcatcagtgcat	2246
Qy	1692	gaaaattgggttttgcgtgtagtaatgatggagttgatgaacaagcaccataattggataa	1751
Db	2247	acctactgaaatcatcagttgatgctaaagtgggcaatggttaccgcgcccaacttacaaca-	2305
Qy	1752	agaacgacttaaaagtggtgctgttgaaattaccacagatagtggttaattaatgctggtaa	1811
Db	2306	-----ctggcgtgaaaaacccagcagcttaacagtgatggcactagtgataaaattag	2357
Qy	1812	tcacaagattaccggacttactbaatggttatagcaaaataccgatcggttaccatccaaca	1871
Db	2358	tgttaaaggttagtggtaagaacatagcttagttaccgccgaacatttggcaagctatct	2417
Qy	1872	gctcaaaagccgaagctcactttaaacgcaggcgatggcatcagtagtaantagtaataa	1931
Db	2418	aaatgaagtcbaatcgaaagctgcagatgcttacaagagcttaccgttaagaagagaaga	2477
Qy	1932	cggggatcagttgtagtagtggtggcaatattaccacccaactataacattagctgtaa	1991
Db	2478	cgatgatgacg-----ccaacgctatacccggtgc	2507
Qy	1992	aaccactaagcttaacagtaatggccaccagtggttaataaatttagttagtaatgc	2051
Db	2508	taaaacacgacaaaaaaatgcggcgagctcagcatctaaacactcaaaaggttaaaacgg	2567
Qy	2052	tcattgataacaaatagcttagttaccgccaaagatttggcagactatcaataaaagtcaa	2111
Db	2568	tctaacgggttgcaccaaaaaagatgtagcttaccttggcttagttagcgaagatagcgg	2627
Qy	2112	tgaacacggctgacagtgcttaccgaagctttaaagtcctaaacacgggttgatataagcaaca	2171
Db	2628	tctgacctatggcacaagaccccttaaacacagatggcttgaactgttaaagataccaacga	2687
Qy	2172	cgccatcacccgtgggttaagatacaaacggcgaagaccttcaacaccttaaacctcaaaag	2231
Db	2688	acaaatccaagtcggtgctaagtggcattaaatttactaatgtgaatggtagtaaatccagg	2747
Qy	2232	tgaacacgggttgaatttaccgaaccaatagagccacaggtacaggttacctttggcaatga	2291
Db	2748	tactggcattgcataaacctgcatta--ccagagataaaaattggcttggctggtgttc	2804
Qy	2292	ccaaagttaatggttcacacgcctaaagctgacggtagcggtagcgatacaaaatggttaatcg	2351

QY	4500	aaatgcggttgcccaaacacccgctgaccttgcaggggtatcagggacacacacgcttaaaaa	4559		Db	6093	tgctttggttggtgggcaataacatcacctgacccgaaagtaactcgttgctttaggttc	6152
Db	5013	aaatgcggttgcccaaacacccgctgaccttgcaggggtatcagggacacacgcttaaaaa	5072		QY	5631	aaactctgccatcagtgacgggacacacacgacacacacgcaaaaaatctgacggcac	5690
QY	4560	actggggagacttgaccatcaaaagtggcaaacagacacacacaaataagctaacggataa	4619		Db	6153	aaactctgccatcagtgacgggacacacacgacacacacgcaaaaaatctgacggcac	6212
Db	5073	actggggagacttgaccatcaaaagtggcaaacagacacacacaaataagctaacggataa	5132		QY	5691	agcaggtacaaacacacacacaggtgcccacaggtacggttaaaggcttgctggacaac	5750
QY	4620	taaatcgggtggttagcaggtactgaggttgcacttgcacttgcacttgcacttgcacttgc	4679		Db	6213	agcaggacaaacacacacacaggtgcccacaggtacggttaaaggcttgctggacaac	6272
Db	5133	taaatcgggtggttagcaggtactgaggttgcacttgcacttgcacttgcacttgcacttgc	5192		QY	5751	ggcgggttggtggttgcctccgtgggtggtggtggtggtggtggtggtggtggtggtggtg	5810
QY	4680	caatctaacagcgttaagtgcaggtgacacacacacacacacacacacacacacacacac	4739		Db	6273	ggcgggttggtggttgcctccgtgggtggtggtggtggtggtggtggtggtggtggtggtg	6332
Db	5193	caatctaacagcgttaagtgcaggtgacacacacacacacacacacacacacacacacac	5252		QY	5811	agcaggtgaggtcaggtgcccac	5870
QY	4740	agacgcaaacggtcaagccaaagcaaacacacacacacacacacacacacacacacacac	4799		Db	6333	agcaggtgaggtcaggtgcccac	6392
Db	5253	agactcaagcgtcaagccaaagcaaacacacacacacacacacacacacacacacacac	5312		QY	5871	cacccaaagcattgcccac	5930
QY	4800	gggtggcaaggttcacgttaattggtggcaagggcacaacacacacacacacacacacac	4859		Db	6393	cacccaaagcattgcccac	6452
Db	5313	gggtggcaaggttcacgttaattggtggcaagggcacaacacacacacacacacacacac	5372		QY	5931	taaaagcgaatgcaggtatttcac	5990
QY	4860	acaacaggttaaacgaagtagcgaacacttgggtggttgccttgccttgccttgccttgc	4910		Db	6453	taagggcgaatgcaggtatttcac	6512
Db	5373	acaacaggttaaacgaagtagcgaacacttgggtggttgccttgccttgccttgccttgc	5432		QY	5991	tcttgccagatccatggttaccgggggtattgcccacacacacacacacacacacacacacac	6050
QY	4911	tgacggcgaatcaggttaaacatctccacacacacacacacacacacacacacacacacac	4970		Db	6513	tcttgccagatccatggttaccgggggtattgcccacacacacacacacacacacacacacac	6072
Db	5433	tgacggcgaatcaggttaaacatctccacacacacacacacacacacacacacacacacac	5492		QY	6051	agtggaagcgtggaagctgctggaataatggtggaataatggtggaataatggtggaataatg	6110
QY	4971	taacccacgtcactcaaacagcagcaggttacttggcgtggaaggttaaacgataaccga	5030		Db	6573	agtggaagcgtggaagctgctggaataatggtggaataatggtggaataatggtggaataatg	6159
Db	5493	taacccacgtcactcaaacagcagcaggttacttggcgtggaaggttaaacgataaccga	5552		QY	6111	cgatacccaagggccatgtagggggcaggtggtgaggttgcactttt	6159
QY	5031	aaaacttgccactggttggttacaagtggcgtggaataagcggcaacgcttaacgggga	5090		Db	6633	cgatacccaagggccatgtagggggcaggtggtgaggttgcacttt	6681
Db	5553	aaaacttgccactggttggttacaagtggcgtggaataagcggcaacgcttaacgggga	5612		RESULT	5		
QY	5091	tttaagcaatgttggttcaaaaccccaaaagatggcagcaaaaaagccctgctgccac	5150		AAF59100	standard; DNA; 6973 BP.		
Db	5613	tttaagcaatgttggttcaaaaccccaaaagatggcagcaaaaaagccctgctgccac	5672		ID	AAF59100		
QY	5151	ttataacgcccaggtcagacacactatgtgacacacacacacacacacacacacacacac	5210		XX	AAF59100;		
Db	5673	ttataacgcccaggtcagacacactatgtgacacacacacacacacacacacacacacac	5732		AC			
QY	5211	ataaatgacaaggtatccgttcttccatgtcaacgagtggaatcaagagcctgtggt	5270		XX	24-APR-2001 (first entry)		
Db	5733	ataaatgacaaggtatccgttcttccatgtcaacgagtggaatcaagagcctgtggt	5792		DT			
QY	5271	acaagggcgaagcagatgggtgaagcccgcccttgcctcagcagcagcagcagcagcagc	5330		XX	M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene SEQ ID NO:1.		
Db	5793	acaagggcgaagcagatgggtgaagcccgcccttgcctcagcagcagcagcagcagcagc	5852		DE	Moraxella catarrhalis strain 4223; major outer membrane protein;		
QY	5331	ccagggcgaagcagatgggtgaagcccgcccttgcctcagcagcagcagcagcagcagc	5390		KW	200kDa outer membrane protein; antibacterial; immunogenic; infection;		
Db	5853	ccagggcgaagcagatgggtgaagcccgcccttgcctcagcagcagcagcagcagcagc	5912		KW	otitis media; detection; ds.		
QY	5391	ccaatccatgcacgtgatacgaacacacacacacacacacacacacacacacacacacac	5450		XX	Moraxella catarrhalis.		
Db	5913	ccaatccatgcacgtgatacgaacacacacacacacacacacacacacacacacacacac	5972		OS	WO200107619-A1.		
QY	5451	tacaggcgaatggtgtagcaggtgaagcacttgggtgcactcggcagcagcagcagcagc	5510		PN	01-FEB-2001.		
Db	5973	tacaggcgaatggtgtagcaggtgaagcacttgggtgcactcggcagcagcagcagcagc	6032		PD	26-JUL-2000; 2000MO-CA00870.		
QY	5511	ggcgtataacaggtacaggtgggtgaataacacacacacacacacacacacacacacac	5570		PF	27-JUL-1999; 99US-0361619.		
Db	6033	ggcgtataacaggtacaggtgggtgaataacacacacacacacacacacacacacacac	6092		XX	(CONN-) CONNAUGHT LAB LTD.		
QY	5571	tgctttggttggtgggaataacatcacccgtgacccgaaagtaactcgttgcttaggttc	5630		PR	Loosmore SM, Sasaki K, Yang Y, Klein MH;		

xx

Example 3; Fig 2A-W; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (Ab), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa gene, which is used CC in the exemplification of the present invention.

xx Sequence 6973 BP; 2265 A; 1555 C; 1533 G; 1620 T; 0 other;

Query Match 66.4%; Score 4090.6; DB 22; Length 6973;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;

Qy 1 atgaatcacatctataaagtcatttttaacaaagccacagggcacatttatggcgtggcg 60
Db 542 atgaatcacatctataaagtcatttttaacaaagccacagggcacatttatggcgtggcg 601
Qy 61 gaatatccaaatcccaagctac-ggggggggtagtggtgtacagggcaagtggcag 119
Db 602 gattacgccaatcccaagctac-ggggggggtagtggtgtacagggcaagtggcag 661
Qy 120 tglacgacttaagctttcccgatttgcggcgctgctgctcgtgactgctgctgag 179
Db 662 tgtatgcaactgagctttcccgatttgcggcgctgctgctcgtgactgctgctgag 721
Qy 180 gctoaatggcagtgcttatgtctcaacaaattactacca---agatgaaattggctcaac 236
Db 722 gctcagtgaggctgttatgtctcaacaaattactaccaatacctatcgtgagca 781
Qy 237 aaacaagataaacaacgctgaagcgatgcctcagcagcagcagtgatgaagcattcgc 296
Db 782 aaacagccaagcgtc-----agcgactgcgaagcgagcgtgagcagcattgc 835
Qy 297 ttttggtagcttcttaaggccaagcgtctcaagctattgctcgttagtggtcacaac 356
Db 836 tatttggaaatgctaacgcacagcggcgttcaagcattcgccatcggtagtagtaata 895
Qy 357 agatccataatagtgtagtaattgtaattgtaggttccacgcgcaaaagtaacagtgccat 416
Db 896 aactgtcaatgggaagcagtttggataagataggtaacgatgtacaggtgtcaagagtc 955
Qy 417 cgcacgtggtgtagtattggtgaaggtgagtgctcgtcgtcgtcgtcgtcgtcgtgata 476
Db 956 cgcacgtggtgtagtattggtgaaggtgagtgctcgtcgtcgtcgtcgtcgtcgtgata 1015
Qy 477 ctatatttgcctaagaatcttgatctgaagaa-----tgaatttcacaa 521
Db 1016 cttaacatttgcgtcagcagtggttaactctaaacatccgaaaggtactgtgataacga 1075
Qy 522 acttattcattggccatgaaatattaaaaaatacaaacctcaacccagtgtaaaatcaa 581
Db 1076 tcttattaacggccatcagattataaagaataacagcgtcgaaggtcaaaagataatgatgaa 1135
Qy 582 atatcagcgcacaagacacaaagcgcagtcactgctcgtggtggagccatgtcatatgc 641
Db 1136 atatagacgcacaacgcagcgcagcgcagtcactgctcgtggtggagccatgtcatatgc 1195
Qy 642 acaggggtcatttttccaacgcttttggatcatatcagcagcagcgtgaagctcctattcctt 701
Db 1196 acaggggtcatttttccaacgcttttggatcatatcagcagcagcgtgaagctcctattcctt 1255

Qy 702 ggcagtaggtttgcgcgcccccaagccacaaaatacttcaatcgtgtgtgttccaatgc 761
Db 1256 ggcagtaggtttgcgcgcccccaagccacaaaatacttcaatcgtgtgttccaatgc 1315
Qy 762 aaaagcacaagcgtttgcagcagcagccattggtggaaaatacttagttaattgggctgc 821
Db 1316 aacatcagctcgttggagcgatagcccttggcaggttactcgtcgtcagctacaggg 1375
Qy 822 aggcgttgcctaggttttggcttcagatccttgataggataataacacagatgccag 881
Db 1376 cagtattgcctaggttcaaggttctgtgtcaactcagatgataataatc---tagacc 1432
Qy 882 tgcctatgtaccactaggttaaaacgttagcagacagctataaaagcccccaggggtga 941
Db 1433 ggcctatacacaataaccagggcactagacccaagtttcaagccac---caataatc 1489
Qy 942 ttctacggatataatttccattgtgataatagtaataataataacagcagctacagcgtaa 1001
Db 1490 gaaggcgggtccacttccattgtgtagtaa-----ctctataaacgttaa 1534
Qy 1002 aatcataatgcgtgcgggttctcgggataccgagtcggtcgaatgtggcacagcttaa 1061
Db 1535 aatcataatgcgtgcgggttctcgggataccgagtcggtcgaatgtggcacagctaga 1594
Qy 1062 atgtgtggagaaactggcta---atcgtaaaattacttttaagggtgagtgtaacataa 1118
Db 1595 agcgttggtagtggtgctaaagcgtagaattacttcttcaggggtgatg-----ataa 1648
Qy 1119 tagcaatagcgtagaagaggtttgggcaatacacttaacttaaggtgagtgacacag 1178
Db 1649 cagtactgaactgaataataggttggataacttaacttaaggtgagtgacacag 1708
Qy 1179 caacgcttaaccgaaagcgtacacatcggtgtgttaacagatggcaa-----tggctcga 1232
Db 1709 caacgcttaaccgaaataataatcggtgtgtgttaaaagagcgtgataatagtggtcga 1768
Qy 1233 agttaactgtctaaagagcgtgactggattgacaggtgtctcc-----gc 1277
Db 1769 agttaactgtctaaacttttaacaaacttctactgaggtgaatacacaactataatgc 1828
Qy 1278 taccacaaaatcccggttagtaataacacaaacacacgcgcagctacaaaagcgtgc 1337
Db 1829 cacaacacagctgaagtaggtagtagtagtagtagtagtagtagtagtagtagtagtag 1888
Qy 1338 tttagacttttagccca-----ataacaggtacaaaacagataaaaacgctctacagcat 1391
Db 1889 tttaacctttaccagcccaatacagcagcagcagcagcagcagcagcagcagcagcagc 1948
Qy 1392 tgaatgagttgaattactaataatgataatgataatgataatgataatgataatgataat 1451
Db 1949 taatgggtgagttactaataatgataatgataatgataatgataatgataatgataatg 2008
Qy 1452 taccacaaaagaaatgt 1511
Db 2009 taccagagataaaattgt 2068
Qy 1512 tcttgacaacgaaagcgttaaaagttggcaacagcagcagcagcagcagcagcagcagc 1571
Db 2069 ttttgataaaaaaacacttaaaagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2128
Qy 1572 taatacaccaactggttaataaacaataccaagtcggtgtgtgtgtgtgtgtgtgtgtgt 1631
Db 2129 tgcaggttaataaaaaagatcagtaacttgcgaaggtagcagtgtagcagtgtagcagtg 2188
Qy 1632 agtcgcttaataatgttgcaataacacagcagcagcagcagcagcagcagcagcagcag 1691
Db 2189 catcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2248
Qy 1692 gaaatgt 1751
Db 2249 acctactgaaatatacagttgtagtgaaggtggaatgttaccgcccccaacttacaaca - 2307

QY	1752	agaaacgacttaaaagtgggtcgttgtagaaattaccacagatagttggtattaatgcgtgtaa	1811
Db	2308	-----tfgggcgtaaaaacccacocagcttaaacagtgatggcactagtataaatattag	2359
QY	1812	tcaagaattaccgacttactactaattggttatagcaaatcccagatcggttaccacaaca	1871
Db	2360	tgtaaaggtagtggtaoagaacaaatagcttagttaccgccgaacatttggcagctatct	2419
QY	1872	gtcaaagcgccaagcctactttaaacgagcgcatggcattcagataattaataataa	1931
Db	2420	aaatgaagtcaatcgaaacggtcgacagtgctctcaaaagctttaccgtttaagaagaaga	2479
QY	1932	cggggatctagttgatagtagtggcaatatbaccaccaactttaacattatagcgtgaa	1991
Db	2480	cgatgatgcg-----cc	2509
QY	1992	aaccactaaacttaaacagtaaatggcacccagtggtgataataaatttagtgttagtaatgc	2051
Db	2510	taaaagacagcaaaaaaattcccgccagtcagcatcttaaaactcaaaaggtaaaaaagg	2569
QY	2052	tcatgataaacaatagcttagttaccgcccaaaagtattggcagactatctaataaagtcaa	2111
Db	2570	tctaacggttgctaccaaaaaagatggtacggttacctttgggcttagccagatagcgg	2629
QY	2112	tgaacggtcgacagtgctctaccaagctttaaagtcctcaaaacggtgataatagaacaaa	2171
Db	2630	tctgaccttggtgcaaaagcaccctaaacaacgagtggtgactggttaaaagatacccaagc	2689
QY	2172	cgccatcacccgtgggtaaagtatacaaacgccaagacacctcaacaccttaaaactcaagg	2231
Db	2690	acaaataccagtcggtgctaaatggcattaaatttactaaatgtaagttagtaatccagg	2749
QY	2232	tgaacacggttgaattatagacaaatagagccacaggttacaggttaccotttggcattga	2291
Db	2750	tactggcattgcaaataccgctgcattat---ccagagataaaattggcttggctggtc	2806
QY	2292	ccaaagttaatggtctccacagcgttaagctgacggtgggttagcgatacaaatggtaatcg	2351
Db	2807	tgatggtgagttgatac-----aaacaaccttactctgtatcaagacaaagcta--ca	2857
QY	2352	attggttattggaagtccttagcgtcgctgacggttaacagcccaaaaaacattattaaag	2411
Db	2858	agttggcaatgtaagattaccaacacitggcattaacgaggtggttaaagcatacaggg	2917
QY	2412	attgtcccaaacctgccttagatttcagcttcagctcaagtgccgcaacatagcactgggcaa	2471
Db	2918	gctgtcccaaacactgcttagatttgcgatacaaatgaagtacgcaacatagaactgggcaa	2977
QY	2472	tacaatcgaaagaaagacaaatcccaacgctgcagcattgtagtgcgttaaatgcagg	2531
Db	2978	tacaatccaaagacaaagacaaatcccaacgctgcagcattgtagtgcgttaaatgcagg	3037
QY	2532	ctttaacctaaaaataatgcaagacaaagacttgcctccacttatgacactgttga	2591
Db	3038	ctttaacctaaaaataatgcaagacaaatcccaacgctgcagcattgtagtgcgttaaatgcagg	3097
QY	2592	ctttatcgatggcaatgccaccacccgacagtaacttatgatgaagccaatcaaacag	2651
Db	3098	ctttggccaatggcaatgccaccacccgacagtaacttatgatgaagccaatcaaacag	3157
QY	2652	taagtggtgatgatgtagtgaatggtgagatgagaaaacattgaaactgacagggcagataatgg	2711
Db	3158	taagtggtgatgatgtagtgaatggtgagatgagaaaacattgaaactgacagggcagataatgg	3217
QY	2712	caagaacaaacttgggttcaaaaaacataaacctgacccaagaaacagtaacttaatggtatgc	2771
Db	3218	caataaaaaacttgggttcaaaaaacccaaactgaaacaaacagtgcttaatggttaatac	3277
QY	2772	aactacat-----ttagtacgcagatgacacatgcccttggtttaaaagcccggttatat	2822
Db	3278	agcaactaaccttaattgtaactcttagtgaagaagtgccctgtttaacgccaaagacat	3337
QY	2823	cgccgggcaactcaaacaccccttagccgaggaatttcacaccacaaaggccagcaaacac	2882

QY	6111	cgataccgaagccatgttagggcgccagttggtgcaggttttcaacttt	6159
Db	6635	cgataccgaagccatgttagggcgccagttggtgcaggttttcaacttt	6683
RESULT 6			
AAF28545			
ID	AAF28545	standard; DNA; 62909 BP.	
XX	AC	AAF28545;	
XX	04-APR-2001	(first entry)	
XX	Genomic fragment #32.		
XX	Genomic library; bacteria; human upper airway; otitis media; sinusitis;		
XX	bronchopulmonary; endocarditis; meningitis; ss.		
XX	Moraxella catarrhalis.		
XX	WO200078968-A2.		
XX	PD	28-DEC-2000.	
XX	16-JUN-2000;	2000WO-US16649.	
XX	18-JUN-1999;	99US-0140121.	
XX	(INCY-) INCYTE GENOMICS INC.		
XX	Lagace RE, Patterson C, Berg KL;		
XX	WPI; 2001-041427/05.		
XX	Genomic library for identifying diagnostic and therapeutic		
XX	compositions, and for identifying virulence factors, regulatory		
XX	elements and drug targets, comprises Moraxella catarrhalis nucleic		
XX	acids -		
XX	Claim 1; Page 293-308; 545pp; English.		
XX	The present invention relates to a Moraxella catarrhalis genomic library		
XX	comprising of a combination of 41 nucleic acid molecules (see		
XX	AAF28514-AAF28554). The library has a number of uses described in the		
XX	specifications. M. catarrhalis (Branhamella catarrhalis) is a large		
XX	aerobic, gram-negative diplococcus, normally found among the bacterial		
XX	flora of human upper airways. M. catarrhalis is known to cause acute,		
XX	localised infections such as otitis media, sinusitis and bronchopulmonary		
XX	infection and life-threatening, systemic diseases including endocarditis		
XX	and meningitis.		
XX	Sequence 62909 BP; 18449 A; 13798 C; 12744 G; 17918 T; 0 other;		
XX	Query Match	64.74; Score 3982; DB 22; Length 62909;	
XX	Best Local Similarity	81.6%; Pred. No. 0;	
XX	Matches 5068; Conservative	0; Mismatches 785; Indels 355; Gaps 26;	
QY	1	atgaatcacatctataaagtcatctttaaacaagccacagggcacattattgcccgtggcg	60
Db	42630	atgaatcacatctataaagtcatctttaaacaagccacagggcacattattgcccgtggcg	42689
QY	61	gaatagccaaatcccacagtagtacggggggggg-tagctgtctcacaggcaagtggcag	119
Db	42690	gaatagccaaatcccacagtagtacggggggggg-tagctgtctcacaggcaagtggcag	42749
QY	120	tgtacgcaacttaagcttggccgtattgcccgtctgcgtctcctcgtgatcggtgcag	179
Db	42750	tgtacgcaacttaagcttggccgtattgcccgtctgcgtctcctcgtgatcggtgcag	42809
QY	180	gctcaatggcagtgcttatgctcaacaataattactaccaagatcgaaattggtcaacaaa	239

[illegible]


```
Qy 5652 cacacagcagggcacacacagccaaataatctgacggcacagcaggtacacacacacacagc 5711
    |||||
Db 48024 cacacagcagggcacacacagccaaataatctgacggcacagcaggtacacacacacacagc 48083
    |||||
Qy 5712 aggtgccacaggtacggttaaaagcttctgctgacaaacggcggttgggtgctccgt 5771
    |||||
Db 48084 aggtgccacaggtacggttaaaagcttctgctgacaaacggcggttgggtgctccgt 48143
    |||||
Qy 5772 ggggtcctcaggtactgaacgcgctatccaaaatgtggcagcaggtgaggtcagtgccac 5831
    |||||
Db 48144 ggggtcctcaggtactgaacgcgctatccaaaatgtggcagcaggtgaggtcagtgccac 48203
    |||||
Qy 5832 cagcaccagatgcgtcaatggtagccaggtgtacaaagccacccaaagcattggccaaagc 5891
    |||||
Db 48204 cagcaccagatgcgtcaatggtagccaggtgtacaaagccacccaaagcattggccaaagc 48263
    |||||
Qy 5892 aaccaatgagcttgaccatgctatccacaaacgaaataaaagccaatgagggatttc 5951
    |||||
Db 48264 aaccaatgagcttgaccatgctatccacaaacgaaataaaagccaatgagggatttc 48323
    |||||
Qy 5952 atcagcagatgcgtgcgtccatgccacagcctacattcctgacagatccatgggttac 6011
    |||||
Db 48324 atcagcagatgcgtgcgtccatgccacagcctacattcctgacagatccatgggttac 48383
    |||||
Qy 6012 cgggggtattgccaccacacacagcgtcgaagtcgcgtggcagtggtgactgtgcgaagctgtc 6071
    |||||
Db 48384 cgggggtattgccaccacacacagcgtcgaagtcgcgtggcagtggtgactgtgcgaagctgtc 48443
    |||||
Qy 6072 ggataatggccaatgggtatttaaaatacaatgttcacggatccacaaagccagtgagg 6131
    |||||
Db 48444 ggaataatggccaatgggtatttaaaatacaatgttcacggatccacaaagccagtgagg 48503
    |||||
Qy 6132 ggcgcagtggtgcaggttttcaacttt 6159
    |||||
Db 48504 ggcgcagtggtgcaggttttcaacttt 48531
    |||||
```

RESULT 7

AAF59101

ID AAF59101 standard; DNA; 5979 BP.

AC AAF59101;

XX AAF59101;

DT 24-APR-2001 (first entry)

XX M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa coding sequence #2.

DE Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection; ds.

XX Moraxella catarrhalis.

OS Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX P-PSDB; AAB69133.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis

XX Example 3; Fig 2A-W; 247pp; English.

PS

xx

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines, (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa coding sequence, which is used in the exemplification of the present invention.

xx SQ Sequence 5979 BP; 1974 A; 1360 C; 1355 G; 1290 T; 0 other;

Query Match 64.1%; Score 3947.8; DB 22; Length 5979;
Best Local Similarity 80.9%; Pred. No. 0;
Matches 4905; Conservative 0; Mismatches 1002; Indels 156; Gaps 20;

Qy 166 gtgatcggtgcgacgctcaatggcagtgcttatgtctcaacaaataactaccacag---atc 222

Db 1 gtgatcggtgcgacgctcaatggcagtgcttatgtctcaacaaataactaccacatc 60

Qy 223 gaaattggtcaacaaacagaataaaacacacgctgaaggcgatgcctagcgacaggt 282

Db 61 gcaattggtgaacaaacacagcagcagcagcgcgc-----aggcactgcgaaggcgaggt 114

Qy 283 gaagcatccattgctttgtagtcttcttaaggacacaggtctcaagctattgtctatc 342

Db 115 gatcgagccattgctattggtgaaatgctaacgcacagcggtgcagccatcgcatc 174

Qy 343 ggtagtgtcaaacacagatccctaataatggttagtaattgtaattgtggttccacgcacaa 402

Db 175 ggtagtgtcaataaaactgtcaatggaagcaggttggaataagataggtagctgctacg 234

Qy 403 ggtaacgagtcacgcgcctcgtggtgatgtattgctgaggtgaggtgagctcgattacc 462

Db 235 ggtcaagagtcacgcgcctcgtggtgatgttaaagggtgtagtgggtgctcgattgcc 294

Qy 463 atcggtagtgatgacttatttgcctaaagaaatctgtatctgaagaa----- 509

Db 295 atcggtagtgatgacttatttgcctaaagaaatctgtatctgaagaaatctgaagaaaggt 354

Qy 510 ---tgaattccaaacttattcattggtccatgaaatattaaaaatacaaacctcaacc 567

Db 355 actctgattaacgacttatttaaacggcctgagtgatgatttaaaagaataacagagctcaag 414

Qy 568 gatggtataaatcaaatatcgacgcacagacagagggcagccagctactgcagtgga 627

Db 415 gataatgagtcaaaatagacgcacacccgcgaagcgacagccagctactgcagtgga 474

Qy 628 gccatgcatatgcacaggggtcattttcccaacgcttttgggtacatacgcaacagctgaa 687

Db 475 gccatgcatatgcacaggggtcattttcccaacgcttttgggtacacgggcaacagctaaa 534

Qy 688 gctgcctattccttgccagtaggtgttgcgcacacagccacaaacaatcttcaatcgct 747

Db 535 agtgcctattccttgccagtaggtgttgcgcacacagccacaaacaatcttcaatcgct 594

Qy 748 gttggttccaatgcaaaagctaaacgcttttcagcgcagcagccattggtggaaatactgta 807

Db 595 attggttctgatgcacacatcagctcgttgggagcgatagcccttgggtgaggtactcgt 654

Qy 808 gtttaattgggtgcagggcgttgccttagttttggttctcagatccttgataggagataat 867

Db 655 gctcagctacagggcagtagtgccttagtcaaggttctgttgcactcagagtgataat 714

Qy 868 aatacagatgccaagtccttatgtaccactaggtataaacggttagcagagcaccagataagcc 927

Thu Sep 13 14:19:05 2001

715	aattc---tagaccggcctacacccaataaccaggcactagaccccaagtcttaaacg	771	1759	gttaaagaagaagacgatgatgacg-----ccaac	1788
716					
717					
718					
719					
720					
721					
722					
723					
724					
725					
726					
727					
728					
729					
730					
731					
732					
733					
734					
735					
736					
737					
738					
739					
740					
741					
742					
743					
744					
745					
746					
747					
748					
749					
750					
751					
752					
753					
754					
755					
756					
757					
758					
759					
760					
761					
762					
763					
764					
765					
766					
767					
768					
769					
770					
771					
772					
773					
774					
775					
776					
777					
778					
779					
780					
781					
782					
783					
784					
785					
786					
787					
788					
789					
790					
791					
792					
793					
794					
795					
796					
797					
798					
799					
800					
801					
802					
803					
804					
805					
806					
807					
808					
809					
810					
811					
812					
813					
814					
815					
816					
817					
818					
819					
820					
821					
822					
823					
824					
825					
826					
827					
828					
829					
830					
831					
832					
833					
834					
835					
836					
837					
838					
839					
840					
841					
842					
843					
844					
845					
846					
847					
848					
849					
850					
851					
852					
853					
854					
855					
856					
857					
858					
859					
860					
861					
862					
863					
864					
865					
866					
867					
868					
869					
870					
871					
872					
873					
874					
875					
876					
877					
878					
879					
880					
881					
882					
883					
884					
885					
886					
887					
888					
889					
890					
891					
892					
893					
894					
895					
896					
897					
898					
899					
900					
901					
902					
903					
904					
905					
906					
907					
908					
909					
910					
911					
912					
913					
914					
915					
916					
917					
918					
919					
920					
921					
922					
923					
924					
925					
926					
927					
928					
929					
930					
931					
932					
933					
934					
935					
936					
937					
938					
939					
940					
941					
942					
943					
944					
945					
946					
947					
948					
949					
950					
951					
952					
953					
954					
955					
956					
957					
958					
959					
960					
961					
962					
963					
964					
965					
966					
967					
968					
969					
970					
971					
972					
973					
974					
975					
976					
977					
978					
979					
980					
981					
982					
983					
984					
985					
986					
987					
988					
989					
990					
991					

Db	5014	gaaGCCaTgacagaataaattgaacaaggatccgcgttctccatgtcaacgatggcaat	5073
Qy	5257	caagagcctgtgtacaaggcgtaacggcattgaactcaagtgcctcaggaagcaactca	5316
Db	5074	caagagcctgtgtacaaggcgtaacggcattgaactcaagtgcctcaggaagcaactca	5133
Qy	5317	gtggcgataggtttccaggcccaaggcagatagtgaaagccgcgtgtgcataaggcagacaa	5376
Db	5134	gtggcgataggtttccaggcccaaggcagatagtgaaagccgcgtgtgcataaggcagacaa	5193
Qy	5377	acccaagcaggcaaccaatccatcgccatcggtgtataacgcacaagccagggcgatcaa	5436
Db	5194	acccaagcaggcaaccaatccatcgccatcggtgtataacgcacaagccagggcgatcaa	5253
Qy	5437	tcctatcgccatcggtatacaggcaatgtggttagcagatgaagcactctggtgcatactggcgac	5496
Db	5254	tcctatcgccatcggtatacaggcaatgtggttagcagatgaagcactctggtgcatactggcgac	5313
Qy	5497	ccaagcactgttaaagctgtataacagttacagttacagttgtgtgtataacaacacagttaccgat	5556
Db	5314	ccaagcactgttaaagctgtataacagttacagttacagttgtgtgtataacaacacagttaccgat	5373
Qy	5557	gccactcaaacccgaatgtcttgggtgtgggcaataacatcacctgacccgaagtaactcg	5616
Db	5374	gccactcaaacccgaatgtcttgggtgtgggcaataacatcacctgacccgaagtaactcg	5433
Qy	5617	gttgcccttaggttcaaaactctgccatcagttgcagttgcacacgcagggcacacaaagccaaa	5676
Db	5434	gttgcccttaggttcaaaactctgccatcagttgcagttgcacacgcagggcacacaaagccaaa	5493
Qy	5677	aaatctgaocgcacagcaggtatacaaccaacacagcaggtgtccacaggttacggttaaaaggc	5736
Db	5494	aaatctgaocgcacagcaggtatacaaccaacacagcaggtgtccacaggttacggttaaaaggc	5553
Qy	5737	tttctgtgacaaacgcggttgggtgcgtctccgttgggtgcctcaggtgtcgaacgcggt	5796
Db	5554	tttctgtgacaaacgcggttgggtgcgtctccgttgggtgcctcaggtgtcgaacgcggt	5613
Qy	5797	atccaaaatgtggcagcaggtgaggttcagttgccaccagccagcaggtcgggtccaatggtagc	5856
Db	5614	atccaaaatgtggcagcaggtgaggttcagttgccaccagccagcaggtcgggtccaatggtagc	5673
Qy	5857	caagttgtacaagccaccccaaaagcattgccacgcgaacccaatgaagttgaacatcatc	5916
Db	5674	caagttgtacaagccaccccaaaagcattgccacgcgaacccaatgaagttgaacatcatc	5733
Qy	5917	caccaaaacgaaaataaagccaatgcaggatttcattcagcgatggcgatggcggtccatg	5976
Db	5734	caccaaaacgaaaataaagccaatgcaggatttcattcagcgatggcgatggcggtccatg	5793
Qy	5977	ccaaagcctacattcctgtggcagatccattggttacggggggtattgccaccacaacaggt	6036
Db	5794	ccaaagcctacattcctgtggcagatccattggttacggggggtattgccaccacaacaggt	5953
Qy	6037	caagggtcggttcagtggtgggacttcgaagcctgcgataatggtccaatgggtattttaa	6096
Db	5854	caagggtcggttcagtggtgggacttcgaagcctgcgataatggtccaatgggtattttaa	5913
Qy	6097	atcaatggttcagccgataaccacaagggcattgtagggggcgaggttggtagcaggttttcac	6156
Db	5914	atcaatggttcagccgataaccacaagggcattgtagggggcgaggttggtagcaggttttcac	5973
Qy	6157	ttt 6159	
Db	5974	ttt 5976	

RESULT	8
AAF59106	
ID	AAF59106 standard; DNA; 6259 BP.
XX	
AC	AAF59106;

XX	24-APR-2001	(first entry)	
DT			
XX	M. catarrhalis M56 200Kda gene in pKS348 SEQ ID NO:12.		
DE			
XX	Moraxella catarrhalis strain Q8; major outer membrane protein;		
XX	200Kda outer membrane protein; antibacterial; immunogenic; infection;		
XX	otitis media; detection; ds.		
XX			
OS	Moraxella catarrhalis.		
XX			
PN	WO200107619-A1.		
XX			
PD	01-FEB-2001.		
XX			
PF	26-JUL-2000; 2000WO-CA00870.		
XX			
PR	27-JUL-1999; 99US-0361619.		
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
XX			
PI	Loonsmore SM, Sasaki K, Yang Y, Klein MH;		
XX			
DR	WPI: 2001-159722/16.		
XX	P-PSDB; AAB69127.		
XX			
PT	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,		
XX	useful in protective vaccines and for diagnosis		
XX			
PT	claim 1; Fig 8A-V; 247pp; English.		
XX			
PS			
XX	The present invention describes an isolated and purified nucleic acid (I)		
CC	that encodes a 200 kda outer membrane protein of Moraxella catarrhalis.		
CC	The 200 kda outer membrane protein (II) has antibacterial activity and		
CC	can be used in vaccines. (II), and its truncated versions, are used as		
CC	immunogenic compositions and vaccines to protect against M. catarrhalis		
CC	infections, particularly otitis media in humans. (II) is also used as		
CC	antigen in immunoassays for detecting specific antibodies (Ab), and to		
CC	generate Ab. (I) are used for recombinant production of (II) and its		
CC	fragments are used as probes for identifying/cloning 200 kda protein		
CC	genes from other strains, and for diagnostic detection of M. catarrhalis.		
CC	(I) makes possible production of large amount of recombinant immunogens.		
CC	Expression of truncated versions of (II) reduces toxicity of the protein		
CC	towards the Escherichia coli host. The present sequence represents the		
CC	M. catarrhalis M56 200Kda gene in pKS348, which is given in the		
CC	exemplification of the present invention.		
XX			
SO	Sequence 6259 BP; 2067 A; 1414 C; 1393 G; 1385 T; 0 other;		

Query Match	Score 3946.8;	DB 22;	Length 6259;
Best Local Similarity	64.1%;		
Best Global Similarity	80.9%;		
Conservative	0;		
Nonconservative	Mismatches	1002;	Indels 156;
	Gapa		Gapa 20;

167	QY	tgatcggcgacgctcaataggcaggtcttatgctcaacaataattactacaag---atccg	223
2	Db	tgatcggcgacgctcaggtggcaggtcttatgctcaaaaaaagataccaacatacgg	61
224	QY	aaattggtcaaacacaaagaataaacacacgcgtgaaagcggtgacctagcgacaggtg	283
62	Db	caattggtgaacaaacccgacagacgctc-----aggcactgccaaaggcggaaggtg	115
284	QY	aagcatccattgtctttggtagtcttttaagggcacaaagctctcgaagtattgctatcg	343
116	Db	atcgaagcatctgctattggtgaaatgctaacggcggtcgaagcctgcgcctcgg	175
344	QY	gtagtgtcaaacaccagatcctaataatggttagtaatggtgaatgtagggtccacagccaaag	403
176	Db	gtagtgtcaataaaactgtcatacgggaagcagtttgataagaatggtaccgagtgctacgg	235
404	QY	gtaacgagtcctatgcctcgtgtggtgtagtattatggctgaaggtgagctcgattgccca	463
236	Db	gtcaagagtcctatgcctcgtgtggtgtagtataaagggttagtggtgagctcgattgccca	295

Qy 464 tcggtagtgatgaacttatatttgcctaagaatctttgatctgaaga----- 509
Db 296 tcggtagtgatgaacttatatttgcctaagaatctttgatctgaaga----- 355
Qy 510 -tgaatttcacaaacttattcatggccatgaataataaaaaatacaaacctcaacg 568
Db 356 ctctgattaaacgacttattacacgccaatgagtaataaaagaatacgaagctcaagg 415
Qy 569 atggttaaaatcaaatatcgacgcaagacgaagcgacagcgccagctactgcagtgagg 628
Db 416 ataatgatgtaaaatatagacgcaaacacgcaagcgacacgccaactgcagtgagg 475
Qy 629 caatgtcatatgcacagggctcatctttccaaacgctttgggtacatacgaacacgctgaag 688
Db 476 caatgtcatatgcacagggctcatctttccaaacgctttgggtacatacgggcaacagctaaaa 535
Qy 689 ctgcctattcttggcagtagctcttgcgcccccaagccacaaaaaatcttcaatcgtg 748
Db 536 gtgcctattcttggcagtgggctctgcgcccccaagcgagggccaatctcaaatcgccta 595
Qy 749 ttggttccaatgcacaaagcctaacgctgtttgcagcgacagccattgttggaatactctag 808
Db 596 ttggttcttgatgcaacatctagctcgtttggagcgatagcccttggtgcaggtactcgtg 655
Qy 809 ttaatttgggtcgaggcgttgccttagttgttgcctcagatccttgcataagggaataa 868
Db 656 ctacgctacagggcagttatgccttagtcaagggtctgttgcctcagagtgataata 715
Qy 869 atacagatgcceagctgctctatgtacacactaggtataaacgttagcagaccaggtataagcca 928
Db 716 attc---tagaccgctctacacacaaataaccaggcactagaccacgaagttcaagcca 772
Qy 929 ccgcgcaggggtgattctacgggataatttccattggtaataagtaataataatagca 988
Db 773 c---caataatcacgaagcggtccaaacttccattgg-----tagtaact 814
Qy 989 gtatcagcggtataaatcatcaatgtcggtgcgggttctcgggatacccgatcggtcaatg 1048
Db 815 ctatacaacgtaaaatcatcaatgtcggtgcagggtgttaataaaacagatcggtcaatg 874
Qy 1049 tggcacagcttaaatgttgagggaactggcta---atcgtaaaattacttttaagggtg 1105
Db 875 tggcacagctagaagcgttggtgaagtgggcgaagagcgtagaataacttttccagggtg 934
Qy 1106 atggtgcaataaatagcaatagcgttagaagaaggtttgggcaataactttaactattaaag 1165
Db 935 atg-----ataacagtagcagctaaaaaatagggttgataataactttaactattaaag 988
Qy 1166 gtatgcacagaccacgcattaaaccgaagctaaacatcggtgtgtgtaa-----cagatg 1219
Db 989 gtggtgcagagaccacgcattaaaccgataataatcggtgtgtgtaaaagaggtcgata 1048
Qy 1220 gcaatggtctgaagaagtaaatctgtcaagagcgtgaactggatt-----ga 1264
Db 1049 atagcgttctgaagaagtaaatctgtcaaaactttaaacatttactgaggtgaatacaa 1108
Qy 1265 ccagtgctctcgctaccacacaaaatcacctgttagtaataaccaacacacacgcgcgagc 1324
Db 1109 ctacattaaatgccacacacacagttaaaggtaggtagtagtagtagtagtagtagtagtag 1168
Qy 1325 tacaagaagcgttggtttgacctttagcc-----caataacagggtaacaaaacagataaaa 1378
Db 1169 tattgagttagtttaacctttaccccgcccaatacacggcagtcacagcaagcaaaa 1228
Qy 1379 ccgtctacagattagtaggttaagttacttaataatgataagtaataagtaataag 1438
Db 1229 ccgtctatggtgttaatgggggtgaagttacttaataatgcaagaacacagcagcaatcg 1288
Qy 1439 gtactactcgtattaccacaaaagaataatggtttgctgggtactaatgatgaggttgatg 1498
Db 1289 gcactactcgtattaccagagataaaatggcttgcctcgagatggtgatgttgatgaaa 1348

Qy 1499 aaagcaaacctcttatcttgacaaacgaaaaagcgtaaaaagttggcaacagcacccttaaacagtg 1558
Db 1349 aacagcaccatatttggtataaaaaaacacttaaaagtggttagttgcaataacatag 1408
Qy 1559 gtactgtactgttaataaacaccactggtataaaaaaacatccaagtgcgtgtaagtga 1618
Db 1409 acaatggcatgtgagcaggtataaaaaagatcagtaacttggccaaaggtagcagtgcta 1468
Qy 1619 ttaaatgtgcacagctcgtctaataaattgttgcaaatcacctcagcaaacagtcggcactgctc 1678
Db 1469 acgatgcggttaccatcgaaacagctcaaaagccgaagcctactttaaacgcagcgctcg 1528
Qy 1679 gtattaccgaagagaaaaattggcttctgtgttactaaatgatgtgaggtgatgaacaagac 1738
Db 1529 gcatacgtgtcacacactactgaaatcagttgatgtgaagagtggaaggtttacgcgcc 1588
Qy 1739 catatttggataaagaacgacttaaaagtggtcggtgtgtgaattaccacagatagtgta 1798
Db 1589 caacttaacaa-----ctggcggtgaaaacacccagcgacttaacagtgatggcacta 1639
Qy 1799 ttaatgctgttaatacacagattaccgacttacttaattggtatagcaaaataccgactgcgg 1858
Db 1640 gtgataaatttagtttaagggttagtggtacgaacaatagcttagttaccgcgcgaacatt 1699
Qy 1859 ttacatcaaacagctcaaaagacgccaagcctactttaaacgcagggcgatggcatacagta 1918
Db 1700 tggcagctctataatgaatgaatcgaacggctgacagtgctctacaaagcttttacog 1759
Qy 1919 ttaatagtaataaacgggactctagttgatgtagtggaataattaccacccaactata 1978
Db 1760 ttaagaagaagacgatgatgacg-----ccaaag 1789
Qy 1979 aatagtagtgtaaaacccactaaagcttaacagtaaatggcaccagtggttaataataattta 2038
Db 1790 ctatcacctggctaaagatacgcacaaaaatcgccgcgcagctagcatcttaaaactca 1849
Qy 2039 gtgttagtaatgctcatgataacaatagcttagttacogcccaaaagatttggcagactatc 2098
Db 1850 aaggtataaaacggttactaactggtctacaaaaaagatgtgacggttactcttgggctta 1909
Qy 2099 taaataaagtcaatgaacgcgctgacagtgctctaccaagctttaaaagtcacaaacgggtg 2158
Db 1910 gccagatagcggctgaccttgccaagaagcaccttaaacacacagatggctgactgctta 1969
Qy 2159 aatagtaacaacgcctacacgtgggttaaagatacaaacgcggaagccttcaacacct 2218
Db 1970 aagataccaacgaacaaatccaagtcggtgctaaaggcattaaattactaattggaatg 2029
Qy 2219 taaaactcaaaaggtgaaaaacggtttaaattattacgaccaatagagccacaggtacagtta 2278
Db 2030 gtagttaatccaggtactggcattgcaaataccgctgcgatta---ccagagataaaatg 2086
Qy 2279 cctttggcattgaccaaagtaatgtctcacacgccttaagctgacggtgggtagcgata 2338
Db 2087 gctttgctggttctgatggcgagttgatac-----aaacaaacctctacttctgataa 2139
Qy 2339 caaatgttaatcgattggtttattgagcaagtccttagcgtgactgacggtgacagcaccacaa 2398
Db 2140 gacaagcta--caagtggcaatgttaagattaccaacactggcattaaacgcaggtggtta 2197
Qy 2399 acatcattaaaggaattgtccccaacactgccttagcattgcaggtcccaagtgccgcgaaca 2458
Db 2198 aagcctacacagggctgtccccaacactgccttagcattgcccagatccaagtagcgcaaca 2257
Qy 2459 tagcactggcgaatacaatcgaagaaaaagacaaatccaaactgcacagcttgcagattgatg 2518
Db 2258 tagaactggcgcaatacaatcccaagacaagacaaatccaaactgcacagcttgcagattta 2317
Qy 2519 tgcataatgcaggtttaaacttaaaaaataatgggaagacacaaagactttgtctccact 2578
Db 2318 tattaataacaggtttaaacttaaaaaataataaacacccattgactttgtctccact 2377
Qy 2579 atgacactgttgactttatcgatggcaatgcccaccacccacagtaacttatgatgaag 2638

[illegible]


```
Qy 5332 caggccaaggcagatggtgaagccgcggttgccatagggcagacaaaccccaagcaggcaac 5391
    |||
Db 6115 caggccaaggcagatggtgaagccgcggttgccatagggcagacaaaccccaagcaggcaac 6174

Qy 5392 caatccatcgcatcggtgataacgcgaagccagcgcgatcaatccatcgcatcggt 5451
    |||
Db 6175 caatccatcgcatcggtgataacgcgaagccagcgcgatcaatccatcgcatcggt 6234

Qy 5452 acaggcaatggtgtagcaggtgaagcactctggtgccatcgcgcccaagcactgttaag 5511
    |||
Db 6235 acaggcaatggtgtagcaggtgaagcactctggtgccatcgcgcccaagcactgttaag 6294

Qy 5512 gctgataacaggttacaggtggtgataaacaacacagtttaccgattgcaactcaaacagat 5571
    |||
Db 6295 gctgataacaggttacaggtggtgataaacaacacagtttaccgattgcaactcaaacagat 6354

Qy 5572 gtctttggtggtggaataacatcacctgacccggtgacgaaagtaactcggttgccttaggttca 5631
    |||
Db 6355 gtctttggtggtggaataacatcacctgacccggtgacgaaagtaactcggttgccttaggttca 6414

Qy 5632 aactctgccatcagtcaggcgacacagcagggcacacaaagccaaaatctgacggcgaca 5691
    |||
Db 6415 aactctgccatcagtcaggcgacacagcagggcacacaaagccaaaatctgacggcgaca 6474

Qy 5692 gcaggtacaacacacacagcaggtgcccacaggtgacggttaaaagcgttgcctggacaacgc 5751
    |||
Db 6475 gcaggtacaacacacacagcaggtgcccacaggtgacggttaaaagcgttgcctggacaacgc 6534

Qy 5752 gcggtggtgcgtctccgtggtggtgctcaggtgctgaacgcgcctatccaaaatgtggca 5811
    |||
Db 6535 gcggtggtgcgtctccgtggtggtgctcaggtgctgaacgcgcctatccaaaatgtggca 6594

Qy 5812 gcaggtgaggtgcagtcgacacagcagcagatgcggtcaatggtgacggtgacaggtgacaaagcc 5871
    |||
Db 6595 gcaggtgaggtgcagtcgacacagcagcagatgcggtcaatggtgacggtgacaggtgacaaagcc 6654

Qy 5872 acccaagcattgcgaacgcaaccaatgagcttgacatcgatctccacaaacgaaat 5931
    |||
Db 6655 acccaagcattgcgaacgcaaccaatgagcttgacatcgatctccacaaacgaaat 6714

Qy 5932 aaagccaatgcagggattcatcagcagatgcggtgctgcctcagtcgacacagcctacatt 5991
    |||
Db 6715 aaagccaatgcagggattcatcagcagatgcggtgctgcctcagtcgacacagcctacatt 6774

Qy 5992 cctggcagatccatggttacgggggtattgcccacccacacacaggtcgaaggtgcggtggca 6051
    |||
Db 6775 cctggcagatccatggttacgggggtattgcccacccacacacaggtcgaaggtgcggtggca 6834

Qy 6052 gtgggactgcgaagctgcgggataatggtcaatggtgatttataaataaatggttcagcc 6111
    |||
Db 6835 gtgggactgcgaagctgcgggataatggtcaatggtgatttataaataaatggttcagcc 6894

Qy 6112 gatacccaagccatgtagggcgcgaggttggtgacggttttcacttt 6159
    |||
Db 6895 gatacccaagccatgtagggcgcgaggttggtgacggttttcacttt 6942
```

RESULT 10

AAF59129

ID AAF59129 standard; DNA; 720 BP.

XX AAF59129;

AC AAF59129;

XX 24-APR-2001 (first entry)

DT 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 200kDa partial nucleotide sequence.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection; ds.

KW Moraxella catarrhalis.

OS Moraxella catarrhalis.

XX

PN WO200107619-Al.

XX 01-FEB-2001.

PD 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPT; 2001-159722/16.

XX P-PSDB; AAB69147.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis

XX Example 14; Fig 17; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

XX The 200 kDa outer membrane protein (II) has antibacterial activity and

XX can be used in vaccines. (II), and its truncated versions, are used as

XX immunogenic compositions and vaccines to protect against M. catarrhalis

XX infections, particularly otitis media in humans. (II) is also used as

XX antigen in immunoassays for detecting specific antibodies (Ab), and to

XX generate Ab. (I) are used for recombinant production of (II) and its

XX fragments are used as probes for identifying/cloning 200 kDa protein

XX genes from other strains, and for diagnostic detection of M. catarrhalis.

XX Expression of truncated versions of (II) reduces toxicity of the protein

XX towards the Escherichia coli host. The present sequence represents a

XX M. catarrhalis strain 4223 200kDa partial nucleotide sequence, which is

XX used in an example from the present invention.

XX Sequence 720 BP; 201 A; 142 C; 140 G; 237 T; 0 other;

Query Match

Best Local Similarity 2.7%; Score 166.2; DB 22; Length 720;

Matches 171; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy 1 atgaatcacatctataaaagtcatcttttaacaaagccacagggcacatttatggcgtggcg 60
    |||
Db 542 atgaatcacatctataaaagtcatcttttaacaaagccacagggcacatttatggcgtggcg 601

Qy 61 gaatatgccaatccacacagtcacgggggggggtagctgtgtacagggcgaagttggcagt 120
    |||
Db 602 gagtacgccaatccacacagtcacgggggggggtagctgtgtacagggcgaagttggcagt 661

Qy 121 gtacgactctaaagctttgcccgtattgcccgcctcctcctcgtcgtatcggtgcagc 179
    |||
Db 662 gtatgactctgagcttgcgcgtattgcccgcctcctcctcgtcgtatcggtgcagc 720
```

RESULT 11

AAA92496

ID AAA92496 standard; DNA; 3030 BP.

XX AAA92496;

AC AAA92496;

XX 17-JAN-2001 (first entry)

DT 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) gene from NTHi strain M407.

XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;

KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;

KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;

KW diagnosis; immunogenic; antigen; ds.

XX Haemophilus influenzae.

OS Haemophilus influenzae.

OY

0v 6157 t 6157

3008 + 3008

RESULT 12

AAA92493 DNA: 3036 BP.

AAA
AC AAA92493;XX
DE
17-JAN-2001 (first entry)

XX *..... adhesin (Hia) gene from NTH1 strain 33.*

XX H1A; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine; anti-Haemophilus influenzae antibody; anti-Haemophilus influenzae
KW non-typable Haemophilus influenzae; anti-inflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicæmia; otitis media;
KW antigen; immunogenic; antigen; ds.

XX Haemophilus influenzae.

XX
PN
W0200055191-A2.

XX
21-SEP-2000.XX
DE
16-MAR-2000: 2000WO-CA00289.XX
10 MAR 1980
99US-0268347.XX
 ... COMMERCIAL ... LTD

XX van der Kleij MH:

XX
XX
03/2000/50

DR WPL; 2000-6166577
DR D-PSDB: AAB23854.

XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -

XX
dc
Claim 1. Fig 18: 275pp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia) gene from the non-typeable Haemophilus influenzae (NTHi) strain 33. Hia genes and proteins have antinflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher molecular weight than a full-length protein.

XX
XX

4 22	22222	142:	DS 21:	Length 3036:
------	-------	------	--------	--------------

Query Match	Score	Indels	Mismatches	Gaps
Best Local Similarity	58.0%	0	190	12
Matches	279	Conservative		

5677 aaatctgacggcacagcaggtagaacaccacacgaggtgccacaggtacgggttcaaggc

WO200055191-A2.
21-SEP-2000.
16-MAR-2000; 2000WO-CA00289.
16-MAR-1999; 99US-0268347.
(CONN-) CONNAUGHT LAB LTD.
Loomsore SM, Yang Y, Klein MH;
WPI; 2000-618897/59.
P-PSDB; AAB23857.
Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Hemophilus influenzae infection
Claim 1; Fig 21; 275pp; English.
The present sequence represents a Haemophilus influenzae adhesin (Hia) gene from the non-typable Haemophilus influenzae (NTHi) strain M407. Hia genes and proteins have anti-inflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher amount of recovery than a full-length protein.
Sequence 3030 BP; 965 A; 596 C; 786 G; 683 T; 0 other;

Query Match	Score	DB 21;	Length 3030;
Best Local Similarity	58.3%		
Best Overall Similarity	58.3%		
Best Conservative	58.3%		
Best Indels	58.3%		
Best Mismatches	58.3%		
Best Gaps	58.3%		

[illegible]


```

Db 2555 aaatctgtagtcgagcgggaaacactacaactgctgtgacaaacgggacggttaaacggc 2614
Qy 5737 ttgtctggacaaaacggcggttggtgctggtctccgtggtgctcaggtgctggaacgcgt 5796
Db 2615 ttgcccgtgcaacggcgacgggtgctggttctctgctggcgcaagcgcggaagaaacgt 2674
Qy 5797 atccaaaatgtggcagcaggtgaagtcagtcgacccacacacgacgagtcggtcaatggtagc 5856
Db 2675 atccaaaacgttggcgagggcgaaatttccgctacttccaccgattgcataaacggcagc 2734
Qy 5857 cagttgtacaagccacccaagcatgccaacgcaacaaatgagctgacacatcgatc 5916
Db 2735 cagttgtatgctggcgaaggggtaacaaac-----ctgtctggacaagtg 2782
Qy 5917 caccaaaacgaaaataaaagccaatgcaagggtatttcacagcagtggtgctccatg 5976
Db 2783 aataaagtggcaaacgtgcaagtcaggtacagcaagtcattagcggcttcacagtta 2842
Qy 5977 ccacaagcctacattctgagcagatccatggttacccggggtattgcccacccaacacggt 6036
Db 2843 ccacaagcctctatgctcaggttaaatcaatggttctatttgcgggaagttagttatcaagt 2902
Qy 6037 caaggtgctggcagtgagtgactgcgaagctgctcgataatggtcgaatgggtattttaa 6096
Db 2903 caaagtggttagctatcggtggttcaagaatttccgataatggaagtgcaagtgatttcgc 2962
Qy 6097 atcaatggttcagcgcatacccaagggcagtgataggggcgagtggtggtgcaggttttcc 6156
Db 2963 ttgtcagggcacaaccaatagccaaggttaaacacagcggttcgacaggtgttggttacag 3022
Qy 6157 t 6157
Db 3023 t 3023

```

RESULT 13

AAA92495
ID AAA92495 standard; DNA; 3354 BP.

XX
AC AAA92495;

XX 17-JAN-2001 (first entry)

XX Haemophilus influenzae adhesin (Hia) gene from NTHi strain K9.

XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen; ds.

XX Haemophilus influenzae.

XX WO200055191-A2.

XX 21-SEP-2000.

XX 16-MAR-2000; 2000WO-CA00289.

XX 16-MAR-1999; 99US-0268347.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Yang Y, Klein MH;

XX WPI; 2000-618897/59.

XX P-PSDB; AAB23858.

XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection

XX Claim 1; Fig 22; 275pp; English.

XX

CC The present sequence represents a Haemophilus influenzae adhesin (Hia)
CC gene from the non-typeable Haemophilus influenzae (NTHi) strain K9.
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.

XX Sequence 3354 BP; 1082 A; 639 C; 866 G; 767 T; 0 other;

Query Match 2.2%; Score 138.2; DB 21; Length 3354;
Best Local Similarity 57.4%; Pred. No. 1.1e-25;
Matches 276; Conservative 0; Mismatches 193; Indels 12; Gaps 1;

Qy 5677 aaatctgacggcagcaggttacacccacacagcaggtgcccacaggttcaggttaaacgc 5736

Db 2864 aaatctgagtgacggcggttaacactacaactgctgcaacacgggttacgttaaacgc 2923

Qy 5737 ttgtctggacaaaacggcggttgctgctccgtggtggtgctcaggtgctgaaacgct 5796

Db 2924 ttgcccgtgcaacggcgacgggtgctgcttctgctggcgcaagcggaagaagact 2983

Qy 5797 atccaaaatgtggcagcaggtgaggtcagtcacacgacacgacgagtcgaatgtagc 5856

Db 2984 atccaaaacgtcgcgagggcgaaatttccgccacttcacgcagtcgattaacggcagc 3043

Qy 5857 cagttgtacaagccacccaaagcattgccaacgcacccaatgagctgaccatgctac 5916

Db 3044 cagttgtatgctg-----tggcaaaaggggtaacaaaatcttctggtgacaagtg 3091

Qy 5917 caccaaaacgaaaataaaagccaatgcagggtatttcacagcagtcgagtcggtccatg 5976

Db 3092 aataaagtggcgaacgtgcagatgcagtcacagcaagtcattagcattcacagtta 3151

Qy 5977 ccacaagcctacattctctgagcagtcctggttacgcgggggtattgcccacccaacgct 6036

Db 3152 ccacaagcctctatccaggttaatacaatggttctctatgctgggaagtagttacaaggt 3211

Qy 6037 caaggtgctggcgagtggtgagtcggaagctgctggaataatggttcaatgggtatttaa 6096

Db 3212 caaaatggttagctatcggtggtatcacgaatttcgcgataatggcgaagtgatttcgc 3271

Qy 6097 atcaatggttcagccgatacccaagccatgtagggcgagtcaggtggtcaggtttcac 6156

Db 3272 ttgtcagggcacaaccaatagccaaggttaaacacagcggttgcagcaggtgtgtaccag 3331

Qy 6157 t 6157

Db 3332 t 3332

RESULT 14

AAA92498

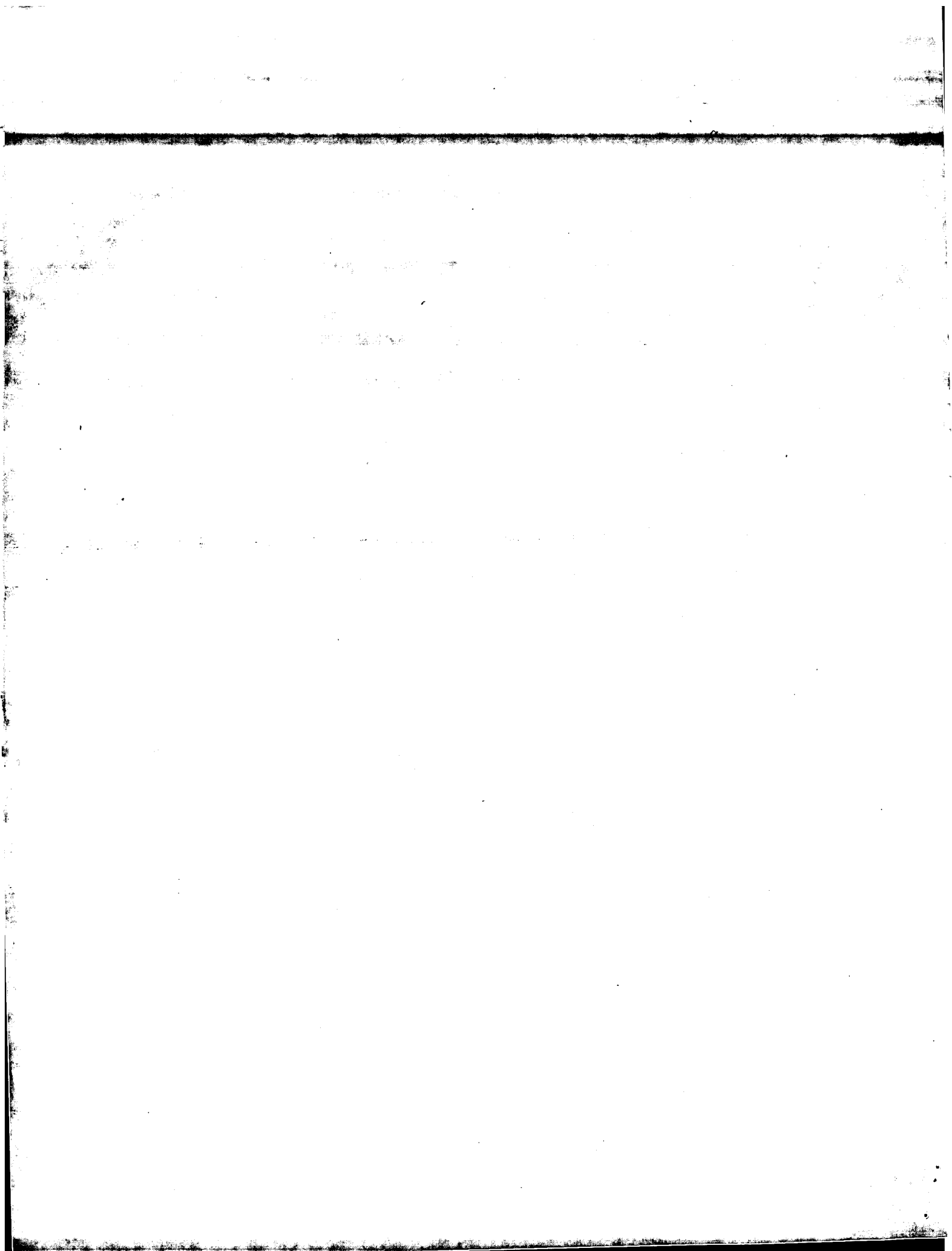
ID AAA92498 standard; DNA; 3342 BP.

XX
AC AAA92498;

XX 17-JAN-2001 (first entry)

XX Haemophilus influenzae adhesin (Hia) gene from NTHi strain K22.

XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:54:07 ; Search time 454.28 Seconds
(without alignments)
2566.627 Million cell updates/sec

Title: US-09-361-619-8
Perfect score: 6159
Sequence: 1 atgaatcacatctataaagt.....ttggtgcagggtttcacttt 6159

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4090.6	66.4	6973	1	US-08-478-370-1
2	4001.4	65.0	9542	4	US-08-968-685A-9
3	84.4	1.4	3294	1	US-08-409-995-1
4	84.4	1.4	3294	3	US-08-685-467-1
5	84.4	1.4	3294	4	US-08-913-942-1
6	83.6	1.4	2037	4	US-08-913-942-14
7	77.2	1.3	7291	4	US-08-913-942-3
8	66	1.1	5738	1	US-08-409-995-3
9	66	1.1	5738	3	US-08-685-467-3
10	60	1.0	60	1	US-08-478-370-4
11	58.4	0.9	1797	4	US-09-377-155-12
12	56.8	0.9	1797	4	US-09-377-155-4
13	55.2	0.9	1770	4	US-09-377-155-18
14	55.2	0.9	1776	4	US-09-377-155-10
15	55.2	0.9	1776	4	US-09-377-155-20
16	55.2	0.9	1779	4	US-09-377-155-3
17	55.2	0.9	2308	4	US-09-377-155-1
18	53.8	0.9	1785	4	US-08-232-463-14
19	53.6	0.9	1785	4	US-09-377-155-6
20	53.6	0.9	1785	4	US-09-377-155-8
21	53.6	0.9	1800	4	US-09-377-155-14
22	52	0.8	1779	4	US-09-377-155-16
23	48.4	0.8	414	2	US-08-630-822A-63
24	48.4	0.8	414	2	US-09-005-069-63
25	48.2	0.8	2277	1	US-08-676-967-2
26	48.2	0.8	2277	1	US-08-676-974-2
27	48.2	0.8	2277	2	US-09-098-487-2

28	45.8	0.7	4765	1	US-08-750-532-8	Sequence 8, Appli
29	43.6	0.7	3489	2	US-08-728-323A-1	Sequence 1, Appli
30	43.6	0.7	32207	2	US-08-770-379-20	Sequence 20, Appli
31	43.6	0.7	32207	4	US-08-757-669A-20	Sequence 20, Appli
32	42.2	0.7	15222	2	US-08-801-898A-23	Sequence 23, Appli
33	42.2	0.7	15222	4	US-08-962-690-12	Sequence 12, Appli
34	42.2	0.7	15223	2	US-08-892-403A-1	Sequence 1, Appli
35	41.4	0.7	273	2	US-08-630-822A-65	Sequence 65, Appli
36	41.4	0.7	273	2	US-09-005-069-65	Sequence 65, Appli
37	40.6	0.7	293	2	US-08-630-822A-71	Sequence 71, Appli
38	40.6	0.7	293	2	US-09-005-069-71	Sequence 71, Appli
39	40	0.6	615	4	US-08-998-416-186	Sequence 186, App
40	39.2	0.6	819	1	US-08-320-161-3	Sequence 3, Appli
41	39.2	0.6	819	3	US-08-642-807A-18	Sequence 18, Appli
42	39.2	0.6	819	4	US-08-455-829-3	Sequence 3, Appli
43	39.2	0.6	1991	4	US-08-235-836C-69	Sequence 69, Appli
44	39.2	0.6	1991	4	US-08-235-836C-77	Sequence 77, Appli
45	39	0.6	1462	3	US-08-961-083-41	Sequence 41, Appli

ALIGNMENTS

RESULT 1
US-08-478-370-1
; Sequence 1, Application US/08478370
; Patent No. 5808024
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,370
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-502
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-478-370-1

Query Match 66.4%; Score 4090.6; DB 1; Length 6973;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 5063; Conservative 0; Mismatches 1009; Indels 157; Gaps 21;
Oy 1 atgaatcacatctataaagtcatcttttaacaaagccacaggcacatttatggcgtggcg 60
|||||

Db	542	ATGAATCACATCATATTAAGTCATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGGCA	601
Qy	61	gaatatgcaaaatccacacagtagac-ggggggggggtagctgtctacaggggcaagttgcag	119
Db	602	GAGTACGCCAATCCACACAGCAGCGGGGGGGTAGCTGTGTACAGGGCAAGTTGGCAG	661
Qy	120	tgtagcactetaagcttggccgattcccgctgcgtgcctcgtgcatcgatcgatgcag	179
Db	662	TGTATGCACTGTGAGCTTTGCCCGTATTTGCCGGCTCGCTGTCTCTCGTATCGGTGCAAC	721
Qy	180	gctcaatggcagtgcttatgctcaacaaatctactacca---agatcgaaatgggtcaaac	236
Db	722	GCTAGTGGCAGTGTCTTATGCTCAAAAAAAGATCAACAACATATCGCAATTTGGTGAACA	781
Qy	237	aaacaagataaaacacacgctgaaagcgatgccttagcgacaggtagaagcatccattgc	296
Db	782	NAACACAGCAAGACGCTC-----AGCAGCTGCCAAGGGCGAGGTGATCGAGCCATTGC	835
Qy	297	ttttgtagtctttctaagggcacaagcgtctcaagctattgcttatcggtagtgctcaaac	356
Db	836	TATTTGGTGAATGCTAACGCACAGGGCGGTCAAGCCATCGCCNTCGGTAGTAGTAATAA	895
Qy	357	agatocctaataaagtagtaagtaagtagtggttccacgcgcaaaaggtaaacaggtccat	416
Db	896	AACTGTCAATGGAAGCAGTTTGGATAGATAGGTACCGATGCTACGGTCAAGAGTCCAT	955
Qy	417	cgccatcggtgtagtattggctgaggggtgtagtgcctcgatgctgcacatcgtagtatga	476
Db	956	CGCATCGGTGGTATGTAAGGGCTAGTGTGTATGCTCGATTTGCCATCGGTAGTATGA	1015
Qy	477	cttatattgcttaagaatcttgatctgaagaa-----tgaatttcacaa	521
Db	1016	CTTACATTTTGTGTATGATGACATGGTAATCTTAACATCCGAAAGGTACTCTGTATTAACGA	1075
Qy	522	acttattcatggccatgaaatatataaaaaaaatacaaacctcaaccgatggtaaaatcaa	581
Db	1076	TCATTATTACGGCCATGCAGTATTAAAGAAAAATAGAAAGCTCAAGAGGATAATGATGTAA	1135
Qy	582	atctgcagcgacaagagcacaagggcagccagtagctactcagtgggagccatgctcatatgc	641
Db	1136	ATATAGACGCACAACCGCAAGCGGACCGCCAGTACTGCAGTGGGAGCCATGTCTATATGC	1195
Qy	642	acaggttcatttttccaaagcctttgggtacatcagcaacagctgaagctgcttattcctt	701
Db	1196	ACAGGGTCAATTTTCCAAAGCCTTTGTATACGGGCAACAGCTAAAGTGGCTATTCTCT	1255
Qy	702	ggcagtaggtgttcgcgcacaaagccacaaacaaatctcaatcgctgttggttccaatgc	761
Db	1256	GGCAGTGGGTCTTCGCCGCCACAGCCGAGGCCCAATCTACAATCGTATTGGTTCTGATGC	1315
Qy	762	aaaagctaacgcggtttgcagcgacagccatttggtagaaatgtagttaatttgggtcg	821
Db	1316	ACATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGCAGGTACTCTGTCGCTACGCTACAGGG	1375
Qy	822	agggcttgocctaggttttggcttcagatcccttgataggataataatacacagatgccag	881
Db	1376	CAGTATTGGCCCTAGCTCAAGGTCTGTGTGTCTACCTACAGAGTATTAATATC---TAGACC	1432
Qy	882	tgccatgtaccactaggtataaaacggttagcagaccagataaaagccaccgcgcagggtaga	941
Db	1433	GGCCTATACACCAATACCCAGGCACCTAGACCCCAAGTTTCAAGCCAC---CAATAATATAC	1489
Qy	942	ttctacggatataattttccattggtataatagtaataataatagcagtagtatcagggcgtaa	1001
Db	1490	GAAGGGGGTCCCATTTTCCATTGGTAGTAA-----CTCTATCAACAGGTAA	1534
Qy	1002	aatcatcaatgctgggtgcgggtttctcgggataccgcagtcggtcgaatgtggcacagcttaa	1061
Db	1535	AATCATCATGTCGGTGCAGGTGTTAATAAATAAACCCGATCGCGTCAATGTGGCAGCAGTAGA	1594
Qy	1062	attgggtggaggaactggcta---atcgtaaaatcttttaagggtgatggtgacaataa	1118
Db	1595	AGCGGTGGTGAAGTGGGCTTAAGAGCGGTAGAAATTTACTTTTCAGGGTGTG-----ATAA	1648

QY	1119	tagcaatagcgtagaagaagcttggcgcaatactttaactattaaagtgatgcacgac	1178
Db	1649	CAGTACTGACGTAAATAATGGTTGGTAAATACTTTAACTATTAAAGGTGGTGCAGAGAC	1708
QY	1179	caacgattacccgaagctaacactcgggtggtgtaacagatggcaa-----cggtctgaa	1232
Db	1709	CAACGCATTAAACCGATATAATATCGGTGGTGGTAAAGAGGCTGATAATAGTGGTCTGAA	1768
QY	1233	acttaaaccttctaagaagctgactggatggaccagtgctccc-----gc	1277
Db	1769	AGTTAAACTTGTATAAACTTTTAAACAATCTTACTTGAGGTGAATACAACTACATTAATATGC	1828
QY	1278	taccacaataatcacgcgttagtaataccaacaacaacacgcgcgagctacaaagcggtag	1337
Db	1829	CACAACACACAGTTAAGGTAGGTAGTAGTAGTACTACAGCTGGAATATTAGTGTATAG	1888
QY	1338	tttgaccttttagccca-----ataacagggtacaaaacagataaaaaccgtctacaagat	1391
Db	1889	TTTTTAACCTTTACCCAGCCCAATACAGGCAGTCAAGACACAAAGAAAACCGCTATGCGCGT	1948
QY	1392	tgatgattgaagtttactaatgataglaaagtagatagcaactaaaggtactactcgatat	1451
Db	1949	TAATGGGGTCAAGTTTACTTAATAATTCAGAAACACACACAGCAATTCGGCTAT	2008
QY	1452	tacaaaaagaaatttgggttctgctgggtactaatgatgaggttgatgaagacaaacctta	1511
Db	2009	TACCAGAGATAAAATTTGGCTTTGCTCGAGATGGTGATTTGATGAAMAAACNAGCCACATA	2068
QY	1512	tcttgacaacgaagaactaaaagtttgcaacagcccttaaacagttggtagctgactgt	1571
Db	2069	TTTGGATATAAAACAACTTTAAAGTGGTGTAGTTGCCAAGGTACCACTGCTACAGTACG	2128
QY	1572	taataacaccacttggtaataaacaacaaacaaatcggtcgctaatggcattaaattggcac	1631
Db	2129	TGCAGGTAATAAAAAAGATCATGTAATCTGCCAAGGTACCACTGCTACAGTACGCTTAC	2188
QY	1632	agtcgctaataattgtcaataacctcagcaacagtcggcactcgctgtattaccagaaga	1691
Db	2189	CATCGAACNGCTCAAAAGCCCAAGCCCTACTTTAAACGGACGGCTGGCGATCAGTGCAC	2248
QY	1692	gaataattggttctggtgtaactaatgatgagttgatgaacagcaccattattggataa	1751
Db	2249	ACCTTACTCAAAATATCAGTTGATGCTAAGTTCAGGTGCCAATGTTACCGCCCAACTTACAACA-	2307
QY	1752	agaaacgacttaagtgggtctgttgtaaaataccacaagatagtggtattaatgcggtaa	1811
Db	2308	-----TTGGCGTGA AAAACCCACCGACGCTTAACAGTGATGGCACATGATGATTAATTTAG	2359
QY	1812	tcacaagattaccgaactactaatggtatagcaaataccgatcggtgtaccatcaaaaa	1871
Db	2360	TGTTAAGGTATGTTGTCGAACATAGCTTAGTTACCGCGCAACNTTTGGCAACAGCTATCT	2419
QY	1872	gctcaagacgcgaagcctactttaaacgcggcgatggcatcagatttaagttaataa	1931
Db	2420	AAATGAAGTCAATTCGAACGGCTTGACAGTGCTTCATAAAGCTTTACCGTTTAAAGAGAAGA	2479
QY	1932	cggggactagttgatagtagtggcaaatataccaccccacttataaacattagcgtgaa	1991
Db	2480	CGATGATGACG-----CCAAACGCTATCACCGCTGGC	2509
QY	1992	aaccacttaagcttaacagtaattggcaccagtggttaataataaatttagtggtagtaatgc	2051
Db	2510	TAAAGATACGACAAAAAATTCGCGCGCAGCTGACATCTTAAACTCAAAAGTTAAAAACGG	2569
QY	2052	tcatgataacaaatagcttagttaccgcccaagatttggcagactactcaataaagaatcaa	2111
Db	2570	TCTAACGGTTGCTACCAAAAAAAGATGGTACGGTTTACCTTTTGGGCTTACGCCAAGATAGCGG	2629
QY	2112	tgaacgcggtgacagtgctctaccaagcgtttaaagtccaaaacggtgataatagacaaa	2171
Db	2630	TCTGACCATTTGGCAAAAAGACACCTTAACAACAGTATGGCTTTGACTGTTAAAGATACCAACGA	2689

Thu Sep 13 14:19:08 2001

4835 AGCCCAAAACCCAGATGGCATTGGCTCAATGAATGTCAATCACTAGTCATTAAACAAGA 4894
4380 acaagttaaatgagcgaataaaagcaagagcatcaatgaagacaacgcctttgttaaaagg 4439
4895 ACAAGTAATGATGCCAATAAAAGCAAGGATCAATGAAGACAAACGCCCTTTGTTAAAGG 4954
4440 acttgaaagacgctcttgatacaacaacaaacacccagcagtaactgtgggtgattt 4499
4955 ACTTGAAAAGCGCTTCATGATAAACAACCAACCAACCCGCCACTAAGTGTGGGTGATTT 5014
4500 aaatgcgcttgcccaaacaccgctgacctttgcaggggatcacaggcacaacagcggtataaaa 4559
5015 AATGCCGTTGCCAAACACCGCTGACCTTTGCAGGGGATACAGGCACAACGGCTTAAAAA 5074
4560 actggcgagactttgaccatcaaaaggtggcaacacagacacaaataagctaaacgataa 4619
5075 ACTGGCGAGACTTTGACCATCAAGAGTGGCAACACAGACACCAATAAGCTAACCGATAA 5134
4620 taacatcggttggttagcaggtactgatggcttcaactgtcaactgtgcaaaagacctaac 4679
5135 TAACATCGGTGTAGCAGGTACTGATGCTTCACTGTCAAACTTGCCAAAGACCTAAC 5194
4680 caatcttaacaggttaagtcaagtgagcaccacaaattgatgaaaaggcatctctttgt 4739
5195 CAATCTTAACAGCGTTAATGCAGGTGGCACCAAAATTTGATGACAAAGGCGTGTCTTTGT 5254
4740 agacgcaaacggtcaagccaaagcaaacacccctgtctgaagtgcgaatgggctggacct 4799
5255 AGACTCAAGCGGTCAAGCCAAAGCAACACCCCTGTCTAAGTGGCAATGGGCTGGACCT 5314
4800 ggggtgcaaggtcaatgaatgtggcaaaagcacaagaataccgaacgctgccaatgt 4859
5315 GGGTGGCAAGGTCTATCATGTAATGTGGGCAAGGCAACAAAGATACCGACGCTGCCAATGT 5374
4860 acaacagttaaacgaagtacgcaactgttggtgtct-----tggtaatgatacgc 4910
5375 ACACAGTTAAACGAAGTACGCAACTTGTGGGTCTTGGTAATGCTGGTAAATGATAACGC 5434
4911 tgacggcaatcaggttaaacattgcgacatcaaaaagaccacaaattcaggttcatc 4970
5435 TGACGGCAATCAGGTAAACATTTGCCGACATCAAAAAGACCCCAATTCAGGTTTCAATC 5494
4971 taacgcactgtcatcaaacgagcagcaggttacttggcggttaaaagttaaacgatacga 5030
5495 TAACGCACGTGTCATAAAGCAGGACGCTACTTGGCGGTAAAGGTAAATACGATACCGA 5090
5031 aaaaactgcaactggtgtgtacaagtggtggtgataaagacgcaacgcaacgcaacgca 5090
5555 AAAACTTGGCCTGCTGGGTATACAGTGGCGTGTAAAGCGCAACGCTAACGCGCA 5614
5091 tttaagcaatgtttggtgcaaaccccaaaagatggcgcaaaaagccctgctgcaac 5150
5615 TTTAAGCAATGTTTGGGTCAAAACCCAAAGATGGCAGCAAAAAGCCCTGCTGCCAC 5674
5151 tttaacgcgcaggttcagacaactatgtgaccaaccccgagagccattgacag 5210
5675 TTATAACCGCGAGGTACAGCAACATTTTACCACCAACCCCGCAGAGCCATTGACAG 5734
5211 aataaatgaacaggtatccgctcttccatgtcaacgagtggcaatcaagagcctggt 5270
5735 AATAAATGAACAGGTATCCGCTTCTCCATGTCAACGATGGCAATCAAGAGCCCTGTGT 5794
5271 acaaggcgtaacggcattgactcaagtgcctcaggaagcactcaagtggcgataggtt 5330
5795 ACAAGGGGTAAACGGCAATTGACATCAAGTGCCTCAGGCAAGCACTCAGTGGCGATAGTTT 5854
5331 ccaggcaagcagatggtgaagcgcgcttgccatagcagacacaaacccaagcagca 5390
5855 CCAGGCCAAGCAGATGTTGAAGCGCGCTTGGCATAGGCAGACAAACCCCAAGCAGCA 5914
5391 ccaatccatgcacatcggtgataacgacacagcagcggtgataatccatcgccatcgg 5450
5915 CCAATCCATCGCCATCGGTGATACGCACAGCCAGCGGATCAATCCATCGCCATCGG 5974

5451 tacaggcaatgtgtagcagtaagcactctgggtgcatcgagaccacaagcactgttaa 5510
5975 TACAGGCAATGTGTAGCAGGTAAAGCACTCTGGTGGCCATCGGCGACCAAGCACTGTAA 6034
5511 ggcctataacagttacagttgggttaaatacaaacacagtttaccgatatccactcaaacga 5570
6035 GGCCTATAACAGTTACAGTGTGGGTAAATAACAACAGTTTACCGATGCCATCAACCCGA 6094
5571 tgcctttggtgggcaataacatcacctgacggaagtaactcgggtgcttagtttc 5630
6095 TGTCTTTGTGTGGGCAATAACATCACCTGACCGAAGTAACCTCGGTGCTTAGGTTT 6154
5631 aaactctgcacatcagtgacaggcacacacagcagggcacacaaagcccaaaaaatctgacgcac 5690
6155 AACTCTGCCATCAGTGCAGGCACACACAGCAGGCACACAAGCCCAAAAAATCTGACGGCAC 6214
5691 agcaggtacaacacacacacagcaggtgcacacaggtacgggttaaaggctttgctggacaac 5750
6215 AGCAGGTACAAACACACACAGCAGGTGCAACCCGTCAGGTTAAAGGCTTTGCTGGACAAC 6274
5751 ggcggttggtgctgctccgtgggtgctcaggtgctgacacgcttatccaaaatgtgac 5810
6275 GCGGTGTGGTGGCTCTCCGTGGGTGCTCAGGTGCTGAACCGCTATCCAAATGTGGC 6334
5811 agcaggtgaggtcagtgccacacacacagcaggtgcggtcaatggtagcagttgtacaaagc 5870
6335 AGCAGGTGAGGTGAGTGCACACAGCAGCAGGTCAGGTGAGTGGTACGACGTTGTACAAAGC 6394
5871 caccacaagcatgtccacacacacacacagcaggtgaggtgacacacacacacacacacac 5930
6395 CACCAAGCATTTGCCAACCAACCAATGAGTTGACCATCGTATCCACCAACCAACCA 6454
5931 taagcacaatgcaggggtttcagcagatggcgatggcgatggcgatggcgatggcgatggcg 5990
6455 TAAGGCCAATCAGGGATTTTCATCAGCGATGGCGATGGCGATGGCGATGGCGATGGCGAT 6514
5991 tcttgccagatccatggttaccgggggtattgcccacacacacacacacacacacacacac 6050
6515 TCTTGGCAGATCCATGTTACCGGGGTATTGCCACCAACCGGTCAAGGTGGCGGTGGC 6574
6051 agtgggactgctgaagctgctgataatgctcaatggtggttatttaaatcaatggttcagc 6110
6575 AGTGGGACTGTGGAAGCTGTGGGATATGGTCAATGGTATGGTATGGTATGGTATGGTATGG 6634
6111 cgatacccaagcctatgtagggcgaggtggtgaggttttccattt 6159
6635 CGNATCCCAAGCCATGATGGGCGCGAGTTGGTGCAGGTTTTCACCTTT 6683

RESULT 2

US-08-968-685A-9
; Sequence 9, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9542 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-968-685A-9

Query Match 65.0%; Score 4001.4; DB 4; Length 9542;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 515; Conservative 0; Mismatches 996; Indels 306; Gaps 21;

Qy 1 atgaatcacatctataaagtcatctttaaaagccacaggacacatttatggccgtggcg 60
Db 218 ATGAATCACATCTATAAGTCAATCTTTAAACAAAGCCACAGCACATTTATGGCGGTGGCG 277

Qy 61 gaatatccaaatccacagtcacggggggtagctgtctacaggccaagtggcagt 120
Db 278 GAATATGCCAATCCACAGCAGCGGGGGGTAGCTGTGTACAGGCCAAGTTGGCAGT 337

Qy 121 gtacgcacttaagctttggccgtattgcgcgctcgtctcctcgtgacgtgcgcagc 180
Db 338 GTACGCACCTCTGAGCTTTGCCCGATTGGCGCGCTCGCTGCTCCTCGTATCGGTGGCAGC 397

Qy 181 ctcaatggcagtgcttatgctcaacaaattactaccagaatgcgaattggccaacaaac 240
Db 398 CTCAATGGCAGTGTATG-----CAGGTATCGGAATTAGTGAAGCAGAC 442

Qy 241 aagataaacacacgcgtgaagcgatgcctcagcagctgaagcattccattgctttt 300
Db 443 GGGGA-----AAGGGGGAGCCAAATGCGCGGTGATTAATCCATTGCTATT 490

Qy 301 ggtagtctttcaaggcacaaggctcctcaagctattgtctatcggtagtgtcaaacacagat 360
Db 491 GGTGATATTGCTCAGGCATTGGCTCTCAATCTATTGCTATCGGTGACACAAATAAGTT 550

Qy 361 cctaataatgtagtaattgtaattgtaggttccacagcccaaggtaacagatccatgcc 420
Db 551 CATAAATTCAAATAAATAGCTAATATAGGTGCCAAAGCCTCAGGTAATGAGTCCATCGCC 610

Qy 421 atcgggtgtagtattggctgagggtagctcctcagctgcattgcctcggtagtgactta 480
Db 611 ATCGGTGGTGTATATTGGCTCTTGGCTATGCCCTCGATTGCCATCGGTAGTGTGACTTA 670

Qy 481 tatttgctaagaattcttgatctgaaga---atgaatttcaaaacttattcattggccat 537
Db 671 TATTGAAAAGGAAACGGTACAGCAATCTCAGAGCTTCTACCTATTATTTCGCGGACAG 730

Qy 538 gaaatattaaaaaaatacaaacctcaaccgatggtaaaat---caaatatcgacgcaca 594
Db 731 AAAGCATTTAAACGATATATACCAACTAGCTGACACTAATCTTCAAAAATATAGACGCACA 790

Qy 595 agagcaaggcgcagccagctacggagggagccatgtcatatgcacaggttcatttt 654
Db 791 CAGCAGAGGACAGCCGAGTACTGAGTGGAGCCATGTCTATATGCAAAAGGGTCATTTT 850

Qy 655 tccaacgctttgtacacacacagctgaagctgcctattccttgccagtaggtcttt 714
Db 851 TCCAAGCGCTTTGGTACACGGGCAACAGCTGAAGTACCTATTCTTGGCAGTGGTCTT 910

Qy 715 gcccccagaaccacaaaataatcttcaatcgtctgtgttccaaatgcaaaagctaacgcg 774
Db 911 ACCGCCACAGCCAAAGCAGCATCTTCAATCGTGTGTTCTTAATGCAACAGCTATCGG 970

Qy 775 ttgcagcagacagccattgggtggaataactgtagttaattgggtcgagggcgttgcccta 834
Db 971 TTTGAGGACAGCCGTTGGTGGAGTACTCAAGTTAATTTGAATCGAGGTATTGGCCCTA 1030

Qy 835 ggttttggtctcagatccttgatagggataataataacagatgcc-----gtgcc 885
Db 1031 GGTTTTGGTCTCAGGTCCTTCAGAAGGATAATGATGTAATGCAAGCAATGTCAGGGCC 1090

Qy 886 tatgtaccactaggtataaacgttagcacacca---gtlaaagccaccgcgcaggggtgat 942
Db 1091 TATGCACAGATGATACACAGCCATACACCGGTATAAAGCCACCTTCAAGAATGGT 1150

Qy 943 tctacgatataatttccattggaataataataataataataataataataataataata 1002
Db 1151 GCTACGGATGATTTTCCATTGGTAAATAGTAAATGGGAATGA---CAGTATCAGGCGTAAA 1207

Qy 1003 atcatcaatgctcggtggtggtctcgggataccggtcgaatggtgcaatggtgcaacagcttaaa 1062
Db 1208 ATCATCAATGTCGTTGCGAGTTCTCGCGATACCGATGCGGTCAATGTGCGACAGCTTAAA 1267

Qy 1063 ttggtggaaggaactggttaactcttaaaattacttttaagggtgatggtgcaataataatagc 1122
Db 1268 GAGCGGTGAGCGCTGGCTAATCGTCAATTTACTTTTAAAGGTGATGATAGCAAT----- 1321

Qy 1123 aatagctagaaaagagtttgggcaataactttaactttaagggtgacagagcaaac 1182
Db 1322 AATAGTAGAAAAGGTTTGGCAAGACTTTAACTATCACAGGTGGTGCACAGACCAGC 1381

Qy 1183 gcattaaaccgaagctaacatcggtggtggaacagatggcaatggtcctgaaagttaaacctt 1242
Db 1382 GCATTAAACCGATATAACATCGGTGCTGCTACAAATGCGGATGGTCTCAAGTTCAACTT 1441

Qy 1243 gctaaaagctgactggaattgaccagtgctcctccgctacacaaataacacgttagtaat 1302
Db 1442 GCTGAAACTTTAAACAGAGCTTTAAATGTTTACCACCTGAAAACCTTAACGCCCAACGAGAAA 1501

Qy 1303 accaacaacaacgcgcgagcta----- 1326
Db 1502 GTTACCGTAGCAAAACCCTTACCACAGATAAAATTTGTTTACCATGATATGAAT 1561

Qy 1327 ----- 1326
Db 1562 GGCAATTGATGAAAGCAAAACCTTTATCTGTATAAGACACTGGCACTTCATGCAGGTGGTCAA 1621

Qy 1327 ----- 1326
Db 1622 AAGATTACCAAACTTACTGCTGGTGTAGTAGACGATGCGGCAACTTATGACAGCTT 1681

Qy 1327 -----caagcgtggtttgaccttttagcccttagcccaataacacaggtaca 1365
Db 1682 AAAAAAGTTAACCAAAACCGCTGAAAGTCTCTACAAACCTTTTACCCTTTAAAAAGGTAGAT 1741

Qy 1366 aacacagataaaacgcgtctacagcaattgattgaagttttactaatgtagtaattagt 1425
Db 1742 AAAAAATGTTAATGATGCTAATGACAGCAAAATCATCACCGTGGGTAAAAATAAACAACCA 1801

Qy 1426 atagcaactaaaggtaactactcgttatcccaaaaaaagaataatggttttctggtggtactaat 1485
Db 1802 GACGCTACTCAAGTCAACACCTTAAACTCAAAAGGTGAAAACGGTGTGATGTTTACAACC 1861

Qy 1486 gatgggtgtgataagcaaaccttattctgacaacgaaa---gctaaaagtggcaac 1542
Db 1862 GAACAAATGGTACAGTTTACCTTTGGGCTTAAACCAAAATACGGTCTGACCGTTGGCAAC 1921

Qy 1543 agcacccctaaacagtggtgactgtactgttaataacacacactggtaataaacaacaaatccaa 1602
Db 1922 AGCACCTTAAACAAGATGGCTTATCTGTAAAAACACCACTAGTAAACAACAATCCAA 1981

Qy 1603 gtcggtgtaatgacattaaatttggccacagtcgctaataatgttgcataatcactcagca 1662


```

Db 6350 AACGAAATAAAGCCAAATCGAGGATTTTCATCAGCGATGGCGATGGCTCCATGCCACAA 6409
Qy 5983 qccatcctcctgagatccatggttaccgggggtattgccaccacacacggtcaaggt 6042
Db 6410 GCTACATCTCTGGCAGATCCATGGTTACCGGGGTATTGCCACCACCAAGCGTCAAGGT 6469
Qy 6043 gcggtgagcagtgctcgaagctgctcgataatggtcaatggttattttaaatacaat 6102
Db 6470 GCGTGGCAGTGGGACTGTGCGAGCTGTGCGATAATGGTCAATGGGTATTTAAATCAAT 6529
Qy 6103 ggttcacgcatccaagcagctagggcggtgagggcggtggtggtggttcttcaattt 6159
Db 6530 GGTTCAGCCGATACCCCAAGCCCATGTAGGGCGGCGAGTGGTGCAGGTTTTCACATT 6586

RESULT 3
US-08-409-995-1
; Sequence 1, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-1

Query Match 1.4%; Score 84.4; DB 1; Length 3294;
Best Local Similarity 50.0%; Pred. No. 8.3e-13;
Matches 245; Conservative 0; Mismatches 236; Indels 9; Gaps 1;

Qy 5677 aaatctgacggcacaggtacacacacacacacaggtggtcaggtttaaaggtc 5736
Db 2803 AAAGCTGACGGTACTGCGGATAAAACCAAGGCGAAGTGAATGATAAAGTTTCTACC 2862
Qy 5737 ttgtgtgacaaacggcggttgggtggttcctcggtggtggtggtggtggtggtggt 5796
Db 2863 GATGAAAACACGTTGTGACGCTTGTGATCAATCAATCAATCAATCAATCAATCAATCA 2922
Qy 5797 atccaaatgtggcagcaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 5856
Db 2923 ATTGACAATGTGGCTAATGCGGATATTTCTCCACTTCCACCGATCGATTAACGGAAGT 2982

```

```

Qy 5857 cagttgtac-----aaagccaccacaaagcattgccaacgcaacaaatgagcttgac 5907
Db 2983 CAGTTGTATGCTGTGCAAAAGGGGTAAACAACCTTGCTGGCAAGTGAATTAATCTTGAG 3042
Qy 5908 catcgtatccacaaacgaataaagccaatgagggatttctcagcagcagcagcagcag 5967
Db 3043 GCGAAAGTGAATAAGTGGCAACCGTGAGATGCGAGGTACAGCAAGTGCATTAGCGGT 3102
Qy 5968 gcgtccatgccacaacccatctcctgagcagatccatggttaccgggggtatttgcacc 6027
Db 3103 TCACAGTTACCAACGCACTATGCGCAGGTAAATGTTGCTATTTCGGGAAGTACT 3162
Qy 6028 cacaagcgtcaaggtcggtggtgagcagtggtgagcagcagcagcagcagcagcagcag 6087
Db 3163 TATCAAGGTCAAATGTTTAGCTATCGGGGTATCAAGATTTCCGATTAATGCAAAAGTG 3222
Qy 6088 gtatttaaatcaatggttcagcgcgatacccaagggccatgtagggcggtggtggtgca 6147
Db 3223 ATTATTGCTGTGTCAGGCACAAAGTATGTAAGTAAACAGGCGTTGCGAGCGGTGT 3282
Qy 6148 ggttttcaact 6157
Db 3283 GGTACCAGT 3292

RESULT 4
US-08-685-467-1
; Sequence 1, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-685-467-1

Query Match 1.4%; Score 84.4; DB 3; Length 3294;

```


Db	2558	CTTTAACT---	GAGCGAACGGCGGAGCGAAGCTTTCACAGTTGATTTAAATGTGGATG	2614
Qy	2483	aaaaagacaaatccaacgcgtgcagcattgatgtgctaaatgcagcgtttaacctaa	2542	
Db	2615	CGAGGAAAAATCCANTGCAGCAAGTATTGAGATGTATTGCGCGCAGGTTTGGATATTC	2674	
Qy	2543	aaaataatggcaaaagcaaaagactttgtctccaactatgcacactgttgactttatcogatg	2602	
Db	2675	AAGGTATGTGTAATAATGTTTGATTATGTAGCGACGATATGACACAGATAAAGCTTTACCGATG	2734	
Qy	2603	gcataatgccaccaccgcacagtaaacctatgataagcagcaatcaaacaccagtaagaatgacgt	2662	
Db	2735	ACAGCACAGGTACAAACACGGTAAAC-----CGTTAACCCAAAACACAGATGGCAAAAGTG	2788	
Qy	2663	atgatgtgaatgtggatgataagaaaaccattgaactgacaggcgataataggcaaa	2714	
Db	2789	CTCAGCGTTAAATTCGGTGGCGAAACATTCCTGTTATCAAAAGACACCAACGGCAAA	2840	

```

Qy 2663 atgatgtatgttgatgagaacaccatgaactgacaggcgataatggcaa 2714
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2789 CTGACGGTAAATACGTCGCGAAACACTTCGTATTCAAAAGACCACAAAGGCAA 2840

RESULT 9
US-08-685-467-3
; sequence 3, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-685-467-3

```

```

; TELE# 910 211299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 5738 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
; DS-08-685-467-3

```

```

; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-685-467-3

Query Match 1.1%; Score 66; DB 3; Length 5738;
Best Local Similarity 49.4%; Pred. No. 1.1e-07;
Matches 233; Conservative 0; Mismatches 230; Indels 9; Gaps

Qy 2243 ttaattacgacccaataagccacaggtacagttacctttggcattgaccaagtaatg 2302
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2378 TTGAGGTGAAAACATCGGAAAGTCAGTGA'ACTTTTAACGATTGGCGGGGAATACACCTACAG 2437

```


Thu Sep 13 14:19:08 2001

```
RESULT 15
US-09-377-155-20
; Sequence 20, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1773)
US-09-377-155-20

Query Match      0.9%   Score 55.2; DB 4; Length 1776;
Best Local Similarity 51.2%; Pred No. 4.7e-05;
Matches 129; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 5906 accatcgtatccacaaacgaaataaagccaatgcaggatttcacgcgatggcga 5965
   || || || || || || || || || || || || || || || || || || || ||
Db 1520 acaaccgatcgacaatgtggacgcgcgcgtgcggcatcgcccaagcgattgcaa 1579
   || || || || || || || || || || || || || || || || || || || ||
QY 5966 tggcgctccatgccacaagcctacattcttcgcagatccatggttacccggggtattgcc 6025
   || || || || || || || || || || || || || || || || || || || ||
Db 1580 ccgcaggctctggttcaggcgtatttgcccggaagagatgatgcgatcggcgcgcga 1639
   || || || || || || || || || || || || || || || || || || || ||
QY 6026 ccacacaagggtcaagtgctggtggcagtggaactgtcgaagctgtcggaataatggtcaat 6085
   || || || || || || || || || || || || || || || || || || || ||
Db 1640 ctatcgcggcgaagcgcggttacgcacatcggtactccagtatcttcgcagcggaatt 1699
   || || || || || || || || || || || || || || || || || || || ||
QY 6086 ggggtatttaaaatcaatgttcagccgatacccaagcccatgtagggcggcagttggtg 6145
   || || || || || || || || || || || || || || || || || || || ||
Db 1700 ggattatcaaaaggcagcgtctccggcaattccgcgcgcaatttcggtgcttccgcacatcg 1759
   || || || || || || || || || || || || || || || || || || || ||
QY 6146 cagggttttcaact 6157
   || || || || || || || || || || || || || || || || || || || ||
Db 1760 tcgggttatcagt 1771
   || || || || || || || || || || || || || || || || || || || ||
```

Search completed: September 12, 2001, 12:57:17
Job time: 1059 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2001, 03:34:16 ; Search time 18125.9 Seconds
(without alignments)
3211.985 Million cell updates/sec

Title: US-09-361-619-8
Perfect score: 6159
Sequence: 1 atgaatcacatcataaagt.....ttggtgcagggttttcacttt 6159

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_est9.*
- 10: gb_est10.*
- 11: gb_est11.*
- 12: gb_est12.*
- 13: gb_est13.*
- 14: gb_est14.*
- 15: gb_est15.*
- 16: gb_est16.*
- 17: gb_est17.*
- 18: gb_est18.*
- 19: gb_est19.*
- 20: gb_est20.*
- 21: gb_est21.*
- 22: gb_est22.*
- 23: gb_est23.*
- 24: gb_est24.*
- 25: gb_est25.*
- 26: gb_est26.*
- 27: gb_est27.*
- 28: gb_est28.*
- 29: gb_est29.*
- 30: gb_est30.*
- 31: gb_est31.*
- 32: gb_est32.*
- 33: gb_est33.*
- 34: gb_est34.*
- 35: gb_est35.*
- 36: gb_est36.*
- 37: gb_est37.*
- 38: gb_est38.*
- 39: gb_est39.*
- 40: gb_est40.*
- 41: gb_est41.*
- 42: gb_est42.*
- 43: gb_est43.*
- 44: gb_est44.*
- 45: gb_est45.*
- 46: gb_est46.*
- 47: gb_est47.*

- 44: em_esthum10.*
- 45: em_esthum11.*
- 46: em_esthum12.*
- 47: em_esthum13.*
- 48: em_esthum14.*
- 49: em_esthum15.*
- 50: em_esthum16.*
- 51: em_esthum17.*
- 52: em_esthum18.*
- 53: em_esthum19.*
- 54: em_esthum20.*
- 55: em_esthum21.*
- 56: em_esthum22.*
- 57: em_esthum23.*
- 58: em_esthum24.*
- 59: em_esthum25.*
- 60: em_esthum26.*
- 61: em_esthum27.*
- 62: em_esthum28.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estin5.*
- 68: em_estom1.*
- 69: em_estom2.*
- 70: em_estov1.*
- 71: em_estov2.*
- 72: em_estpl1.*
- 73: em_estpl2.*
- 74: em_estpl3.*
- 75: em_estpl4.*
- 76: em_estpl5.*
- 77: em_estpl6.*
- 78: em_estpl7.*
- 79: em_estpl8.*
- 80: em_estpl9.*
- 81: em_estpl10.*
- 82: em_estro1.*
- 83: em_estro2.*
- 84: em_estro3.*
- 85: em_estro4.*
- 86: em_estro5.*
- 87: em_estro6.*
- 88: em_estro7.*
- 89: em_estro8.*
- 90: em_estro9.*
- 91: em_estro10.*
- 92: em_estro11.*
- 93: em_estro12.*
- 94: em_estro13.*
- 95: em_estro14.*
- 96: em_estro15.*
- 97: em_estro16.*
- 98: em_estro17.*
- 99: em_estro18.*
- 100: em_estro19.*
- 101: em_estro20.*
- 102: gb_est25.*
- 103: gb_est26.*
- 104: gb_est27.*
- 105: gb_est28.*
- 106: gb_est29.*
- 107: gb_est30.*
- 108: gb_est31.*
- 109: gb_est32.*
- 110: gb_est41.*
- 111: gb_est42.*
- 112: gb_est43.*
- 113: gb_est44.*
- 114: gb_est45.*
- 115: gb_est46.*
- 116: gb_est47.*

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	83	1.3	806	221	CNS04AEE
c 2	77.4	1.3	641	236	AQ946120 Tetraodon
c 3	71.2	1.2	700	235	AQ940248 Sheared D
c 4	67.4	1.1	729	236	AQ945618 Sheared D
c 5	66.8	1.1	500	107	AU086536
c 6	65.8	1.1	621	236	AQ941683
c 7	62.4	1.0	469	107	AU076358
c 8	62	1.0	541	221	CNS04KUS
c 9	58.2	0.9	827	220	CNS02156
c 10	56.8	0.9	520	231	AQ656286
c 11	56.8	0.9	854	228	AQ411842
c 12	56.2	0.9	442	107	AU053378
c 13	56.2	0.9	989	220	CNS02HA4
c 14	55.8	0.9	500	107	AU088475
c 15	55.2	0.9	590	107	AU039138
c 16	54.6	0.9	401	231	AQ639568
c 17	54.2	0.9	252	156	C93354
c 18	54.2	0.9	363	156	C92585
c 19	54.2	0.9	816	245	A5335744
c 20	53.8	0.9	743	239	AZ196612
c 21	53.2	0.9	560	231	AQ658275
c 22	53	0.9	500	107	AU087658
c 23	53	0.9	835	221	CNS046DH
c 24	52.8	0.9	595	240	AZ216659
c 25	52.8	0.9	935	239	AZ202424
c 26	52.6	0.9	538	218	AA550663
c 27	52.6	0.9	1042	221	CNS03YTF
c 28	52.4	0.9	581	231	AQ658242
c 29	52	0.8	739	239	AZ196614
c 30	52	0.8	989	220	CNS02HA4
c 31	51.8	0.8	240	107	AU073656
c 32	51.8	0.8	500	107	AU088092
c 33	51.4	0.8	444	231	AQ643512
c 34	51.4	0.8	810	239	AZ199472
c 35	51.4	0.8	856	156	C25694
c 36	51.2	0.8	443	258	PT015E15U
c 37	51.2	0.8	500	156	C93360
c 38	51.2	0.8	639	231	AQ656218
c 39	50.8	0.8	410	107	AU037570
c 40	50.8	0.8	660	239	AZ212542
c 41	50.4	0.8	488	258	PT003A15R
c 42	50.4	0.8	702	236	AQ946396
c 43	50.2	0.8	407	231	AQ644537
c 44	49.8	0.8	302	256	B73937
c 45	49.8	0.8	457	156	C90465

ALIGNMENTS

RESULT 1	CNS04AEE/c	806 bp	DNA	21-MAY-2000
LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone			
DEFINITION	095G24 of library G from Tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL281759			
VERSION	AL281759.1			
KEYWORDS	GSS; genome survey sequence.			
SOURCE	Tetraodon nigroviridis			
ORGANISM	Tetraodon nigroviridis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.			
AUTHORS	1 (bases 1 to 806) Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,			

TITLE	Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
JOURNAL	Freshwater pufferfish Tetraodon nigroviridis
REFERENCE	Unpublished 2 (bases 1 to 806)
TITLE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Saurin,W., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 806)
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES	Location/Qualifiers 1..806 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="095G24" /note="Genoscope sequence ID : COBG095BD12LP1-end : T7"
BASE COUNT	247 a 87 c 99 g 333 t 40 others
ORIGIN	
Query Match	1.3%; Score 83; DB 221; Length 806;
Best Local Similarity	43.5%; Pred. No. 6.5e-11;
Matches	273; Conservative 19; Mismatches 335; Indels 0; Gaps 0;
Qy	1407 tactaatgatagtaagtagcaactaaagggtactactcgtattaccacaaagaaat 1466
Db	795 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAA 736
Qy	1467 tgggttctggtactaatgatggagtgtaggaagcaaaccttcttgcaacgaaaa 1526
Db	735 TAATAATAGTAAATATATATATAATAATAATAGTACTACTACTACTACTATAATAATA 676
Qy	1527 gctaaaagtggcaacagcacccctaaacagtggtagctgttgaataacaccactgg 1586
Db	675 TAGTAATAATGATAATAATAATAATAATAGTAGTACTACTACTACTACTATAATAATAA 616
Qy	1587 taataaacacacccaagtcggtgctaatggcattaaattggccacagtcgctataatgt 1646
Db	615 MAAMAATAATGATAATAATAATAATAATAATAATAATAATAATAATGTTGTAATAATG 556
Qy	1647 tgcataatccctcagcaacagtcggtgctgctgtattaccgaagagaaaattggtttgc 1706
Db	555 MSSTATTATATACMAAHMACMAAMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCM 496
Qy	1707 tgggtactaatgatggagtgtagcaacagcacccatttggataaagacgacttaagt 1766
Db	495 TAATAATAATGTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 436
Qy	1767 gggctgtttgaaattaccacagatagtggtatttaactcgtggtatcaacagattaccgg 1826
Db	435 TAATAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 376
Qy	1827 acttactaatggtatgacaaatcaccgctgctggttaccatcaaacagctcacaagcgcaa 1886
Db	375 TATTAATAGTAAAT 316
Qy	1887 gctactttaacagcaggcgtggtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1946
Db	315 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 256
Qy	1947 tagtagtgccgaatattaccaccccccaactataacattagcgtgaaaaaccactaagcttaa 2006

Db 255 TAATGATATCATAGTAATAATAATAATAATAATAATAATGCTAATAGTAACGCTATTA 196

QY 2007 cagtaatggcaccagtggttaataataa 2033

Db 195 TGATAGTAGTGACAACTAATAACACTGA 169

RESULT 2

LOCUS AQ946120/c 641 bp DNA GSS 27-JAN-2000

DEFINITION Sheared DNA-46J23.TF Sheared DNA Trypanosoma brucei genomic clone

ACCESSION AQ946120

VERSION AQ946120.1 GI:6769385

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 641)

AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Doneison, J., Fraser, C. and Adams, M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: Sheared DNA-46J23.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.

Seq primer: M13-Reverse

Class: shotgun.

Location/Qualifiers

1. 641

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db_xref="taxon:5691"

/clone="Sheared DNA-46J23"

/clone_lib="Sheared DNA"

/note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT 213 a 35 c 10 g 383 t

ORIGIN

Query Match 1.3%; Score 77.4; DB 236; Length 641;

Best Local Similarity 46.3%; Pred. No. 1.9e-09;

Matches 255; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

QY 1555 agtggtagctgactgttaataacacactggtaataacaaatccaagtcggtgctaat 1614

Db 638 ACTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 579

QY 1615 ggcattaaatttggccacagtcgctataatgttggcaaatcaacctcagcaacagtcggcact 1674

Db 578 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 519

QY 1675 gctcgattaccgaagagaaattggttttctggtactactaatgatggagttgatgaaca 1734

Db 518 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 459

QY 1735 gcaccataatttggataaagaacgacttaagtggttcgttgaaattaccacacagatagt 1794

Db 458 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 399

QY 1795 ggtatttaagctggttaatacacaagattaccggaacttaagtggttatagcaaatcccgat 1854

Db 398 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 339

QY 1855 gcggttaccatcaaacagctcaagacgcgaacccctactttaaacccagcgatggcagtc 1914

Db 338 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 279

QY 1915 agtatttaagtaataacggggaatcctagttgtagtagtggtggaatattaccaccccaact 1974

Db 278 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 219

QY 1975 tatacattagcgtgaac 2034

Db 218 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 159

QY 2035 tttagttagtagtgcctcgtatgataacaaatagcttagttaccgcgaagatttggcagac 2094

Db 158 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 99

QY 2095 tatctaaataa 2105

Db 98 GATAATAATAA 88

RESULT 3

LOCUS AQ940248 700 bp DNA GSS 27-JAN-2000

DEFINITION Sheared DNA-42E21.TF Sheared DNA Trypanosoma brucei genomic clone

ACCESSION AQ940248

VERSION AQ940248.1 GI:6763513

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 700)

AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Doneison, J., Fraser, C. and Adams, M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: Sheared DNA-42E21.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.

Seq primer: M13-Forward

Class: shotgun.

Location/Qualifiers

1. 700

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db_xref="taxon:5691"

/clone="Sheared DNA-42E21"

FEATURES

source

AU086536	500 bp	mRNA	EST	27-JAN-2001
LOCUS	AU086536	Sugano Malaria cDNA library	Plasmodium falciparum	cDNA
DEFINITION	clone XPFN2783,	mRNA sequence.		
ACCESSION	AU086536			
VERSION	AU086536.1	GI:12388677		
KEYWORDS	EST,			
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	1 (bases 1 to 500)			
TITLE	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.			
JOURNAL	FULL-malaria: a database for a full-length enriched			
MEDLINE	from human malaria parasite, Plasmodium falciparum			
COMMENT	Nucleic Acids Res. 29 (1), 70-71 (2001)			
	Contact: Junichi Watanabe			
	Institute of Medical Science			
	The University of Tokyo, Department of Parasitology			
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan			
	Tel: 81-3-5449-5378			
	Fax: 81-3-5449-5410			
	Email: jwatanab@nagane.ims.u-tokyo.ac.jp			
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., and Sugano			
	S. Construction and characterization of a full length-enriched and			
	5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).			

```

FEATURES
source
Location/Qualifiers
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone_lib="XPFn2783"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
278 a 34 a 142 t 3 others
BASE COUNT

```

Query Match	1.1%;	Score 66.8;	DB 107;	Length 500;
Best Local Similarity	47.9%;	Pred. No. 1.2e-06;		
Matches 101, Conservative		0: Mismatches 208;	Indels 0;	Gaps 0;

QY	1714	aatgatggaggttgatgaacaaggcaccaattttgataaaagacgacttaagctgcgt	1773
Db	86	AATGATAAATAATGCAAATTATTACCGTATCAATAATAAGAAAATAANAATAANGAA	145
QY	1774	gttgaataattaccacagatatgttgtattaatcgtgtaaatacagaattaccggacctact	1833
Db	146	AATFAAAAAATAATGAATAATAATAATAATAACAACAATAATAATAACAATAATAAT	205
QY	1834	aatggtatagaaaataccgatcggttcaccaaccaagcagtccaagacgccgaagcctact	1893
Db	206	AANTAATAAACNNTAATAATAATAATAATAACANTATAATAATAGCAACAATAATAAT	265
QY	1894	ttaaacgcaggcgatggcatcgactagattaatagtaataacgcgggtcttagttgtagtagt	1953
Db	266	AGCACAAATAATAATAGTAACACAAATAATAATAGTACACAAACAATAATAATAATAAT	325
QY	1954	ggcaatattaccaccoccaaactttaacattagcgtgaiaaacccactaaagcttaaacagtaat	2013
Db	326	AATAATAATAATAGTAACCAATAATAATAATAATAATAATAAGTAACCAACAATAATAANTAANTAATAAC	385
QY	2014	ggcacagtggttaataataaatttagttagtaatgctcatgataaccaatagcttagtt	2073
Db	386	AANTAACAATAATAATAATAATAATAATAAGTACACATAATAATAATAGTAACAATAATAAT	445
QY	2074	accgccaaagattggcagactatctaaataaagctcaat	2112
Db	446	AATAATAATCATAGTACCCTAAATTACTAATAATAATGATAAT	484

RESULT
AQ941683
6

LOCUS	AQ941683	621 bp	DNA	GSS	27-JAN-2000
DEFINITION	Sheared DNA-43D2.TR Sheared DNA Trypanosoma brucei genomic clone				
	Sheared DNA-43D2, DNA sequence.				
ACCESSION	AQ941683				
VERSION	AQ941683.1				
KEYWORDS	GI:6764948				
SOURCE	GSS.				
ORGANISM	Trypanosoma brucei.				
	Trypanosoma brucei				
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.				
REFERENCE	1 (bases 1 to 621)				
AUTHORS	El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.				
TITLE	Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSSs: Sheared DNA-43D2.TF				
	Contact: Najib M. El-Sayed				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: nelsayed@tigr.org				
	Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/ .				

FEATURES
SOURCE

Class: snocgation/Qualifiers

1. .521

organism="Trypanosoma brucei"
/strain="TREU927/4 CUTCat 10.1"
/db_xref="taxon:5691"
/clone_lib="Sheared DNA-43d2"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site:1:
Institute for Genomic Research
Genomic DNA isolated from a C
Trypanosoma brucei (TREU927/4
sheared to give a tight size d
The v + i method used for the
described in detail in Smith,
(making small insert libraries
sequencing projects. In Genom
Approach, eds. M. Vaudin and E
Press, 1999)." 78 g 210 t

318 a 15 C

BASE COUNT	318 a	15 c	78 g	210 t
ORIGIN				

Query Match	1.1%	Score 65.8;	DB 236;	Length 621;
Best Local Similarity	50.5%	Pred. No. 2.5e-06;		
Matches 160; Conservative	0;	Mismatches 157;	Indels	0; Gaps 0;

QY	1748	ataaagaacgacttaagtggttgaaattacacagatagcggattaaatgctg	1807
Db	272	ATAATAAACGAGTAATAATAATAATAATAATAATAATAATAATAATA	331
QY	1808	gtaatcacgaattaccggacttaactaatggtatagaacaataccagtcggtaccctca	1867
Db	332	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	391
QY	1868	aacagctcaaaacacgaagcctctttaaacgcgagcgatggcatcagttataatagta	1927
Db	392	GTGTGTGTATACGAATAATTAATAAAGACGACTTAATAATAATAATAATAATA	451
QY	1928	ataacgggagctatgttgatagtagtggoaatattaccaccccaacttatacaacttagcg	1987
Db	452	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	511


```

LOCUS      CNS02156      827 bp      DNA      GSS      12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ORI end of clone
            224F10 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL176451
VERSION    AL176451.1 GI:7814508
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 827)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis.
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 827)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
JOURNAL   Unpublished
REFERENCE  3 (bases 1 to 827)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT   This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES   Location/Qualifiers
            source            1..827
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="224F10"
            /clone_lib="c"
            /note="Genoscope sequence ID : C0AG224DC05SP1-end :
            PUC-ORI"
BASE COUNT 368 a 203 c 106 g 121 t 29 others
ORIGIN
Query Match 0.9%; Score 58.2; DB 220; Length 827;
Best Local Similarity 45.2%; Pred. No. 0.0003;
Matches 182; Conservative 17; Mismatches 201; Indels 3; Gaps 1;
QY 1860 taccatcaaacagctcaagaagcgaagccttactttaaacgagcgatggcatcagtat 1919
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 211 TAACAACAATATACACACACACACACACACACACACACACACACACACACACAA 270
QY 1920 taatgtaataaacggggatctagttagtagtgaggcaataattaccaccccaacttata 1979
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 271 CAACAACAACAACAATAACAACAACAATAACAGCSATAACAACAATAACACGATACAA 330
QY 1980 cattagcgtgaaccacttaagcttaacagtaatggcaccagtggttaataaatttag 2039
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 CAATACACACACACACACACACACACACACACACACACACACACACACACACGVA 390
QY 2040 tggtagtaatgctatgataca----atagcttagttaccgcccaagattggcagacta 2096
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391 TAACAACAATAACAATAACAATACAGTGATAACAACAATAACAACAATAACAACAATRA 450
QY 2097 tctaaataaagtcgaatgaacggctgacagtgtctaccaggtttaaagtcctaaacgg 2156
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 451 CAGCGATACACACATGATGACGCGATGACACACACACACACACACACACACACACAG 510
QY 2157 tgataatagcaaacgcgcacatccggtgggtaagatacaaacggaagccttcaaacac 2216
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 511 CGATAACACGRATAACAACAATRACAGTGATATACACAATRACAGCGATGACACACAATRA 570
QY 2217 cttaaaactcaaggtgaacacggtgtgttaattattacgaccaat 2259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 CAACAATAACAGTGATRACACAACAATGACAGTGATRACACAAT 613

RESULT 10
LOCUS   AQ656286      520 bp      DNA      GSS      23-JUN-1999
DEFINITION Sheared DNA-27K23.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-27K23, DNA sequence.
ACCESSION AQ656286
VERSION    AQ656286.1 GI:5164054
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei.
ORGANISM   Trypanosoma brucei.
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE  1 (bases 1 to 520)
AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
            Fraser,C. and Adams,M.
TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: Sheared DNA-27K23.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tldb/mdb/tldb/.
            Seq primer: M13-Reverse
            Class: Shotgun.
            Location/Qualifiers
            source            1..520
            /organism="Trypanosoma brucei"
            /strain="TREU927/4 GUTat 10.1"
            /db_xref="taxon:5691"
            /clone="Sheared DNA-27K23"
            /clone_lib="Sheared DNA"
            /note="Vector: pUC18; Site_1: SmaI; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
            sheared to give a tight size distribution (approx 2 kb).
            The v + i method used for the library construction is
            described in detail in Smith, H.O. and Venter, J.C.
            (Making small insert libraries for whole genome shotgun
            sequencing projects. In Genome Sequencing: A Practical
            Approach, eds. M. Vaudin and B. Borell, Oxford University
            Press, 1999)."
BASE COUNT 263 a 44 c 79 g 134 t
ORIGIN
Query Match 0.9%; Score 56.8; DB 231; Length 520;
Best Local Similarity 50.4%; Pred. No. 0.0006;
Matches 139; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 1797 tattaatgctggtaatacacaagattaccggacttactaatggtatagcaaataccgatgc 1856
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 60
QY 1857 ggttaccatcaaacagctcaaacgacgcctttaaagcgcagggcgatggcgcag 1916
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 120

```


Qy	1836	tggtatagcaaataccgatacggtgttaccatcaaacagctcaaaagacgcgaagcctacttt	1895
Db	457	TCCATTAACCTAGTAATCCCTACTGGTACTGGTACTACTTAATCTACTTAATCTACTAC	516
Qy	1896	aaacgcagcggatggtgcatcagatattaatagtaataacgagggatctctagtgtgatatgtg	1955
Db	517	TACTACTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	576
Qy	1956	caatattaccaccccaactataacattagcgtgaaacacactaaagcttaacagtaattg	2015
Db	577	TATTAATCTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	636
Qy	2016	caccagtggaataataaaatttagttagtaagtcgatgatacaaatagcttagttac	2075
Db	637	TACTAATAGTAATAGTGTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	696
RESULT - 12			
LOCUS	AU053378	442 bp	EST
DEFINITION	AU053378 Dictyostelium discoideum SL (H.Urushi-hara) Dictyostelium		28-APR-1999
ACCESSION	AU053378		
VERSION	AU053378.1	GI:4701860	
KEYWORDS	EST.		
SOURCE	Dictyostelium discoideum.		
ORGANISM	Dictyostelium discoideum		
REFERENCE	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.		
AUTHORS	1 (bases 1 to 442)		
TITLE	Yoshino,R., Morio,R. and Tanaka,Y.		
JOURNAL	Developmental CDNA in Dictyostelium discoideum		
COMMENT	Unpublished (1997)		
	Contact: Hideko Urushihara		
	Institute of Biological Sciences		
	University of Tsukuba		
	3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan		
	Email: d402hutsakura.cc.tsukuba.ac.jp		
	PROJECT = Dictyostelium discoideum cDNA project in Japan.		
FEATURES	Location/Qualifiers		
source	1..442		
	/organism="Dictyostelium discoideum"		
	/strain="AX4"		
	/db_xref="taxon:44689"		
	/clone_lib="SLI507"		
	/dev_stage="slug"		
BASE COUNT	238 a 40 C 28 g 135 t 1 others		
ORIGIN			
Query Match	0.9%; Score 56.2; DB 107; Length 442;		
Best Local Similarity	54.6%; Pred. No. 0.00082;		
Matches 112;	Conservative 0; Mismatches 93; Indels 0; Gaps		
Qy	1912	atcagttattatagtaataacgagggatctctagtgtgatatgtgtagtggaataattaccaccca	1971
Db	197	AACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	256
Qy	1972	actataacattagcgtgaaacacactaaagcttaacagtaattgaccagtggtgataat	2031
Db	257	AATA	316
Qy	2032	aaatttagttagtaagtcctcatgatacaaatagcttagttaccgcccaagattggca	2091
Db	317	AATA	376
Qy	2092	gactatcataataaagcgaatgaa	2116
Db	377	TCCCTTGAAAAAAGAAAAA	401

RESULT 15

AU039138
LOCUS AU039138 590 bp mRNA EST 29-MAR-1999
DEFINITION AU039138 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSM436, mRNA sequence.
ACCESSION AU039138
VERSION AU039138.1 GI:3985891
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 590)
AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402huesakura.cc.tsukuba.ac.jp
PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
source
1..590
Location/Qualifiers
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:4689"
/clone="SSM436"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 279 a 78 c 39 g 194 t
ORIGIN

Query Match 0.9%; Score 55.2; DB 107; Length 590;
Best Local Similarity 49.0%; Pred. No. 0.0017;
Matches 147; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 1366 aaacagataaaacccgtctacagcattgagtgattgagtttactaatgtagtagtagt 1425
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 65

Qy 1426 atgcaactaaagtgactactcgattaccacaaagaaattggttctggtactaat 1485
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 AATAATAACATATACCAATAATAATAATAATAATAATAATAATAATAATAATA 125

Qy 1486 gatggagttgatgaagcaaaccttctctgacacgaaagctaaaagttggcaacgc 1545
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 AATAACAATAGTATAATAATAATAACAATAATAACAATAATAATAATAATAATA 185

Qy 1546 accctaagcagtggtagctgactgttaataacacacactggttaataacaaatccaagtc 1605
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 AATAATAATAACACAATAATAATAATAATAATAATAATAATAATAATAATAATA 245

Qy 1606 ggtgctaattggcattaaattggcacagtcgctgctaataattgttgcaaataccctcagcaaca 1665
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 246 TATTCAACTAGTCAAAATAGTGTATATATATATATATATATTCGTCACACGCAACATCATCA 305

Search completed: September 13, 2001, 03:34:48
Job time: 53710 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:37:02 ; Search time 78.73 Seconds
(without alignments)
1580.859 Million cell updates/sec

Title: US-09-361-619-9
Perfect score: 10356
Sequence: 1 MNHLYKVFNKATGTFMAVA.....NGSADTQCHVGAAGVGFHF 2053

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10356	100.0	2053	22 AAB69135	M. catarrhalis str
2	7675	74.1	2047	22 AAB69134	M. catarrhalis str
3	7408	71.5	1992	22 AAB69133	M. catarrhalis str
4	7405	71.5	1992	17 AAW04505	Moraxella 200 kDa
5	7405	71.5	1992	22 AAB69137	M. catarrhalis M56
6	7219.5	69.7	2314	22 AAB69136	M. catarrhalis les
7	1206	11.6	2353	17 AAR99393	Haemophilus adhesi
8	1205	11.6	2411	21 AAB23860	Haemophilus influe
9	1021	9.9	1104	21 AAB23856	Haemophilus influe
10	1021	9.9	1104	21 AAB23859	Haemophilus influe
11	994	9.6	1004	21 AAB23857	Haemophilus influe

12	952	9.2	1002	21 AAB23854	Haemophilus influe
13	809	7.8	2039	19 AAW56322	Haemophilus paraga
14	779.5	7.5	2042	19 AAW56319	Haemophilus influe
15	710	6.9	1094	21 AAB23858	Haemophilus influe
16	690	6.7	1098	17 AAR99392	Haemophilus adhesi
17	583.5	5.6	679	17 AAR99394	Haemophilus adhesi
18	583.5	5.6	679	21 AAB23855	Haemophilus influe
19	553	5.3	2514	21 AAY75097	Neisseria meningit
20	549.5	5.3	1601	18 AAW30292	Non-typeable Haemo
21	526.5	5.1	2383	21 AAB15945	E. coli proliferat
22	513	5.0	1598	18 AAW30291	Non-typeable Haemo
23	512.5	4.9	2599	21 AAY75098	Neisseria meningit
24	512	4.9	1529	14 AAR41732	High molecular wei
25	506	4.9	2048	21 AAY75096	Neisseria gonorrh
26	484.5	4.7	2893	19 AAW98828	H. pylori GHP0 148
27	484.5	4.7	2893	19 AAW71556	Helicobacter polyp
28	468	4.5	1477	14 AAR41724	High molecular wei
29	462	4.5	1477	18 AAW30294	Non-typeable Haemo
30	459	4.4	1338	14 AAR41731	High molecular wei
31	458	4.4	1477	14 AAR41728	High molecular wei
32	457	4.4	1536	15 AAR63505	Haemophilus high m
33	457	4.4	1536	21 AAB01846	Haemophilus influe
34	456	4.4	1536	14 AAR41723	High molecular wei
35	456	4.4	1536	18 AAW30293	Non-typeable Haemo
36	454.5	4.4	1536	14 AAR41725	High molecular wei
37	452	4.4	1477	15 AAR63506	Haemophilus high m
38	451	4.4	1222	21 AAB01830	H. Influenzae stra
39	451	4.4	1228	21 AAB01828	Haemophilus influe
40	451	4.4	1477	21 AAB01848	Haemophilus influe
41	451	4.4	1638	20 AAY00138	Enterococcus faeca
42	451	4.4	1638	20 AAY00140	Enterococcus faeca
43	451	4.4	1638	20 AAY00142	Enterococcus faeca
44	449	4.3	2902	22 AAB46351	H. pylori HPN165 p
45	440	4.2	1612	19 AAW65088	R. prowazekii S-la

ALIGNMENTS

RESULT 1
AAB69135
ID AAB69135 standard; Protein; 2053 AA.
XX AAB69135;
XX
XX 24-APR-2001 (first entry)
XX
XX M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.
XX
XX Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
XX Moraxella catarrhalis.
XX
XX WO200107619-A1.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-CA00870.
XX
XX 27-JUL-1999; 99US-0361619.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX WPI; 2001-159722/16.
XX N-PSDB; AAF59104.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis

ps Claim 1; Fig 4A-V; 247pp; English.

xx The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 08 200kDa protein, which is given in the exemplification of the present invention.

xx Sequence 2053 AA;

Query Match 100.0%; Score 10356; DB 22; Length 2053;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2053; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNHIIYVFNKATGTPMAVAEYAKSHSTGGSCATGQVGSVRTLSFARIALAVLIVIGAT 60
DB 1 mnhlykviinfkatgtfmaevaeayakshstggscatgqvgsvrtlsfariaalavlivigat 60

QY 61 LNSGSAVAAQITTTIEIGQTNKINNTLKGDALATGEASTAFGSLSKAQSQAIAIGSRVDP 120
DB 61 lngsayaqqitttkieigqtnkintlkgdalatgeastafgslskaqsqaiaigsvrxdp 120

QY 121 PNNGSNGNVSASHAGNESIAIGDGVLAEGDASTAGSDLLYLPKNLDLKNRPHKLHGHE 180
DB 121 pnngsngnvsashagnesiaigdvlaegdastagsdillylpknldlknrphklhghne 180

QY 181 ILKKIOTSTGKTKYRTRAGHASTAVGAMSAQAQHFNAFTGATATAPAAVSLAVGLAA 240
DB 181 ilkkiofstgktkyrtraghastavgamsaqaqhfnafgtatataapaavslavglaa 240

QY 241 QATQSSIAVGSNAKANAFATAGTGNVNLGRGVALFGSGILDRDNTDASAVVPLG 300
DB 241 qatqssiavgsnakanafaatagtnvnlgrvgalfgsgildrdntdasayvplg 300

QY 301 KTLADQYKATRQGDSTDFESIGNSNNSIRRKIINVAGSRDSDAVNVAQLKLVEEEL 360
DB 301 ktladqykatrqgdstdfesignsnnsirrkiiinvagssrdtdavnvaqlklveela 360

QY 361 NRKITFKGDDNNSNSVERGLGNTLTITKGAQTNALTEANIGVVTGNGLKVKLAKELTG 420
DB 361 nrkitfkdgddnnsnsverglgntltitkgdaqtnalteanigvvtgnglkvklakeltg 420

QY 421 LTSVSATNKITVNTNNNAELQSGLTFSPITGKTDKTVISIDGLKFTNDSNSTATKG 480
DB 421 ltsvsatnkityvntnnnaelqsgltfspitgktdkvtvisidglkftndsnstatk 480

QY 481 TTRITKIKKIFAGTNDGDESKPYLDNEKLVGNSTLNSGLFTVNTTGNKQIQVGANGI 540
DB 481 ttritkikkifagtndgdeskpyldneklvgnstlnsglftvnttgnkqiqvgangi 540

QY 541 KPATVANNVANTSATVGTARTEEKIFAGTNDGDEAPYLDKERLKVGRVEITTDGSI 600
DB 541 kpatvannvantSATVGTARTEEKIFAGTNDGDEAPYLDKERLKVGRVEITTDGSI 600

QY 601 NAGNHKITGLTNGIANTDAVTKQLDKAPTLINAGDGIISNSNNGDLVDSSGNTTPTYN 660
DB 601 nagnhkitgltnGIANTDAVTKQLDKAPTLINAGDGIISNSNNGDLVDSSGNTTPTYN 660

QY 661 ISVKYTKLNSGNTSGNNKFSVSNADHNSLVTAKDLADYLNKVNEDASALPSEKVVONG 720
DB 661 isvkttklnsngntsgnnkfsvsnadhnslvtakdladylnkvnedasalpsekvvong 720

QY 721 NSNNATTVCKDNTGKTFNTKLKGENGVNITTRATGTVTFGIDQSNGLTTPKLVGSDT 780
DB 721 nsnnatvckdntgktnfntklkngenvnitrattgtvtfgidqsnnglttppklvgsdt 780

QY 781 NGNRLVIEQVPSADGNSNTNIIKGLSPTLPSTASPSGRNIALGNTIEEKDKSNAASIDV 840
DB 781 ngnrlvieqvpSadgnsntniiKGLSPTLPSTASPSGRNIALGntleekdkSnaasidv 840

QY 841 LNAFNLKNGKDKDFVSTVDTDFIDGNATTATVYDEANQTSKVAYDVNVDKTIET 900
DB 841 lnaFnlknngkdkdfvstvtDfIdgnattAtvYdeanQtskvaydvNvdktiet 900

QY 901 GDNGKKQLGVKTKIKLTETSPNGNATFTSDDHALVKASDIAGNLNTLAEETHFKGTAN 960
DB 901 gdngkkqlgvKtkIKLTetSPngnatFTsDDHALvkASDIAGnlNTLAEethfkGTan 960

QY 961 TALQTFVTKAVDENKADDTNATVCGKDTSGKVNTLKLKGNGLDKTDKDCQTVFGIN 1020
DB 961 talqtfvtkavdenKadDTNAtVcgkDTsgKvNtLklKgnGLdkTDKdcQTVfGIN 1020

QY 1021 TQSLKAGDSTTLNNGLSIKNTASNEQIQVGADGVKFAFNVNNGVVGAGIDGTRTRRDE 1080
DB 1021 tqslkagdstTLnnglsIKntasneQiqvgadgvKfAFNVNngVVGagIdgtrtrDe 1080

QY 1081 IGFTGTNGSLDKSKPHLSKGINAGGKKITNIOSETAKNSHDVATGGKTYDLKTELENK 1140
DB 1081 igftgtngslDKSKphLSKGINagGKKITnIOsetAKNShdVatGGkTYdlKteLenk 1140

QY 1141 ISSPATAQNSLHFSVADQCGNNFTVSNPYSSYDTSKTSDDVITFAGENGITTKVNGGV 1200
DB 1141 issPataQnSLHfSVadQCGnnftVsnPySSyDtsKtsDDvITfagengITtkVngGV 1200

QY 1201 RVGIDQTKGLTPPKLTVGNNGKGIIVINSQNGQNTITGLSNTLANVTNDKSGSVRTTQGN 1260
DB 1201 rvgidqtkglTPPKLtvGNngKgiIVInsqngQntITglSntlanVtNdKsgsvrtTqGn 1260

QY 1261 IKKEDKTRAASIVDVLISAGNLOGNEAVDFVSTYDTVNFANGNTTAKVYDDTSKTS 1320
DB 1261 iKKedKtraASiVdVLISagNLoGneAVdfVstYdtVnFangntTAKvYddTSkts 1320

QY 1321 KVVYDVNVDDTTEVKKLLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIAVHLNT 1380
DB 1321 kvvydvNVddTTEvKKllGVkTTTLtStgtGankFalsnOatGDalVKasDIaVhlnt 1380

QY 1381 LSGDITQAKGASOANNSAGYVDAGNKVIYDSTDNKYQAQNDGTVDKTKVAKDKLVAQ 1440
DB 1381 lsgdiTqakGASoANnsAGyVdAGnkViYdSTdnKYQAQndGTVdkTKvAKDKlVAq 1440

QY 1441 AQTDPGTLAQMNVKSVINKEQVNDANKKQGINEDNAFVKGLEKASDNKTKNAATVGD 1500
DB 1441 aqtpdgtLaqMnvKsvINkeQvNdAnkKqGINedNaFvKgleKASdnKtKnaATvGD 1500

QY 1501 NAVAQPLTFAGDTGTAKKLGCTITIKGGQDTNKLTDNNIGVVAGTDGFTVKLAKDLT 1560
DB 1501 navaqplTFagDTGTaKKlGctITIKggQDTnKLtDnnIGvvAGtdGfTVkLaKDLt 1560

QY 1561 NUNSVNAGGKTKIDEXGISFVDANGAKANTPVLSEANGDLJGCKVLSNVGKGTQDAANV 1620
DB 1561 nUnsvNagGkTKIdEXGISfVdANGaKantPVLseANGdlJgCKvLSNVgKGTQdaANv 1620

QY 1621 QQLNEVRNLLGLGNDADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGKGNNDTEKLA 1680
DB 1621 qqLNEvrNllGLgNdADgnQvNIADiKKdPNsGSSnRTViKagTVlGKgNNDteKLa 1680

QY 1681 TGGVQGVQDKDGNANGDLSNVWVKTKQDKGSKKALLATYNAAGQTNVVTNNPAEADIRNE 1740
DB 1681 tggvQGVQdkDgNANGdlSNvWVKtKQDKgSKKallATyNaAGqTNvVTnnPaEADIRne 1740

QY 1741 QGTRFHVNDGNGEPVVGQNGIDSSASCKSHVAICFOAKADGEEAAVATGROTQAGNQSI 1800
DB 1741 qgtrFHVndGngEPvVGQngIDSSasCKSHvAICfoAKADgEEAAvATgROTqAGnQSi 1800

QY 1801 AIGDNAQATGDQSIATGCTGNVAVAGKHSGAIGDPSSTVKADNSYSVGNNOFTDATQTDVFG 1860

Db 1801 aignhaqatgdsiaigtgnvvagkhsgaigdpstcvkxadsysvgnnnqfdatcqtavfg 1860
Qy 1861 VGNITVTESNSVALGNSAISAGTHAGTQAKKSDGTAGTGTATTTAGATGTVKGFAGQTAVG 1920
Db 1861 vgnnitvesnvalgnsaisagthagtqakksdgtagtttttagatgtvkgfagqtavg 1920
Qy 1921 AVSVGASGAERRIQWAAAGEVSATSDAVNGSQLYKATQSTANATNELDHRHONENKAN 1980
Db 1921 avsvgasgaerriqwnaagevsatsdtdavngsqlykatqstlanatnelidhrhonenkan 1980
Qy 1981 AGISSAMAMASMPQAYIPGRSMVTGGTATHNGOCAVAGLSKLSDNGOWEKFINGSADTQ 2040
Db 1981 agissamamaspqayipgrsmvtggtathngocavagvlgsklsdngqvwfkingsadtq 2040
Qy 2041 GHVGAAYGAGFHF 2053
Db 2041 ghvgaavgagfhf 2053
RESULT 2
AAB69134
ID AAB69134 standard; Protein; 2047 AA.
XX
AC AAB69134;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.
XX
KW Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
XX W0200107619-Al.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-CA00870.
XX
XX 27-JUL-1999; 99US-0361619.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
XX WPI: 2001-159722/16.
XX
XX N-PSDB; AAF59102, AAF59103.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis -
XX
XX Claim 1; Fig 3A-W; 247pp; English.
XX
XX The present invention describes an isolated and purified nucleic acid (I)
XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
XX The 200 kDa outer membrane protein (II) has antibacterial activity and
XX can be used in vaccines. (II), and its truncated versions, are used as
XX immunogenic compositions and vaccines to protect against M. catarrhalis
XX infections, particularly otitis media in humans. (II) is also used as
XX antigen in immunoassays for detecting specific antibodies (Ab), and to
XX generate Ab. (I) are used for recombinant production of (II) and its
XX fragments are used as probes for identifying/cloning 200 kDa protein
XX genes from other strains, and for diagnostic detection of M. catarrhalis.
XX (I) makes possible production of large amount of recombinant immunogens.
XX Expression of truncated versions of (II) reduces toxicity of the protein
XX towards the Escherichia coli host. The present sequence represents the
XX M. catarrhalis strain 4223 genomic 200kDa protein, which is given in the
XX exemplification of the present invention.
XX
XX Sequence 2047 AA;

Query Match 74.1%; Score 7675; DB 22; Length 2047;
Best Local Similarity 74.8%; Pred. No. 0;
Matches 1604; Conservative 134; Mismatches 219; Indels 186; Gaps 28;
Qy 1 MNHYKVFENKATGTFMAVEYAKSHSTGGSCATGQGVSVTLTSPFARIAALAVLIGAT 60
Db 1 mhykvvfnkatgtfmaaveyakshstggscatgqgvsvctlsfariaalavligat 60
Qy 61 LNSGAYAOQIITK-LEIGOTNKINNTLKGDALATCASTAFCSLSKAOGSOAIATGSKVP 119
Db 61 lsgsayakdkkhaigeqnprsr--gtakadralaiaigenanaggggalaigssnk 118
Qy 120 DPNGSNGNVGSHAKGNSIAIGGDVLAEGDASIAIGSDDLXL-----PKNLDLKN 170
Db 119 tvngssldkigtatgesiaiggdvksasdasiaigsddhlldghgnpkpbg-tlin 177
Qy 171 EFHKLHGHHEILKKTQSTDGKIKYRRTAQAQGHASTAVGAMSYAOGHFSNAGTATARA 230
Db 178 d---linghavlkeirsskdnvdyrrttasghastavgamsyagqghfsnagftrataks 234
Qy 231 AVSLAVGLAAOATKQSSIAVGSNAKANAFATAIGTNTVNLGRGVALGFGSOILDRNN 290
Db 235 ayslavglataegdstiaigsdatcsslgaialagatraqlggsiaiggsvvvtqsdnn 294
Qy 291 TDASAYVPLGKTLDQYKATRGDSTDFISGNSNNNSIRRKIINVGAGSRDTRAVNV 350
Db 295 s-rpaytpntqalpkfqatnntkagpl-sig-----snskrkiinvagvgnktdavnv 347
Qy 351 AOLKLVLELA-NRKITFGDGNNSNVERGLNLTITKGDQTNALTEANTGVV--TDG 407
Db 348 agleavvkwakeritfg--gddnstdvkigldntltikggaetnaldnnlgyvkeadn 405
Qy 408 NGLVKVLAKELTGLTSV-----SATNKITVSTNNNAELQSGGLTFES-PIGTGK-TDKT 460
Db 406 sgikvlaktlnnltevnnttlnattvkvgsssstaelldsltfcpntgsgstkt 465
Qy 461 VYSIDGLFTNDSNSIATKGTTRITKKGIFAGTNGVDDESKPYLDNEKLKVGNSLTNSG 520
Db 466 vygvngvktfn----- 477
Qy 521 SLTVNNTGNKQIQVGANGIKFATVANNVNTSATVGTARITEEKIGFAGTNGVDEQAP 580
Db 478 -----aettaaigttrtdrkigfardgdvdekqap 508
Qy 581 YLDKERLKVGRVEITDSSGINAGNHKITCLTNGINTDAVTIKOLKDAKPTLNAGDGTISI 640
Db 509 yldkkqlkvgsvaltdngidagnkkslnlakgssandavtleqikaakptlnagagisv 568
Qy 641 NSNNGDLVDSSGNITPTYNISVKTTKLNSNGTSGNNKFSVSNHNNNSLVTAKLADLYL 700
Db 569 tptelsvdaksgnvtcptyngvkteinsdgt--dkfsvkgsgtnnslvtaehlasyl 626
Qy 701 NKVNETADSAALPSFKVQNGDSEN-NAITVGKDT--NGKTFNTLKLKGENVNTNRTATG 757
Db 627 nevnrtaosaiqsfvkeedddanaivakttnagavsilkknglgtvatkk-dg 685
Qy 758 TVTFIDQSGNLGTTTPKLVGSD-----TNGN-----RL 785
Db 686 tvtfisgdsdgtlkgstinnngdltvkdtneqiqvgangikftvnsgnsgpvglantari 745
Qy 786 VIEQV--PSADG-----NSTKNITKGLSPTLPSPASPSG 817
Db 746 trdkigfagsgdvdtntkpyldqdklqgvnkvitntginagkaltgtsptpsiadqs 805
Qy 818 RNIALGNITIEKDKSNASIDDLVNLNAGFNKLNKGDKDFVSTYDVFIDGNATATATVY 877
Db 806 rnieigntiqdkdknaasindilntgnfnknnpidfvstydvdvfangnattatvch 865
Qy 878 DEANQTSKVAYDVNVDEKTIETLTGDKKKQLGKVKIKLTETSTNGN-ATTF--STDDHHA 934
Db 866 dtanktskvvydvndvdtthitgtddnkkglgvkttklntksangntatnfnvnsdeda 925

QY 935 LVKASDIAGNLNLTAEIHTTKGTANTALQOTFTVKKVVDENDKADDTNATTVGKDGTSGRV 994
Db 926 lvnaklaenlntlakeinttktgadtatqftvkvdennadadanaaltvgknanqgv 985
QY 995 NTLLKAGKNGLDIKTKDQGVTFGINTQSLKAGDSTLLNNGLSIKNTASNEQIQVGAD 1054
Db 986 ntltlkgnglnlktkngtvtfginttsglkagks-tindgglsiknptseqiqvgad 1044
QY 1055 GVRKAMV-NGVVVAGIDGTRTRTRREIGFTGNGSLDKSKPHLSKDGINAGGKKTINTQ 1113
Db 1045 gvkfakvnnngvvgagldgttrtrtrdeifgtngslskphlskdglnagggkktintq 1104
QY 1114 SGETAKSHDAVNGGKTYLDLTKLENNKISSTAKTAAONSLEHFSVADEQGNNTVSNPYSS 1173
Db 1105 sgelagshadvcgklyldlktelenkistaktagnshelhsfsvadeqgnntfvsnpys 1164
QY 1174 YDTSKTSDDVITFAGENGITTKVANKGVVRVIGIDQTKLTGTPKLTGNNNGKGINVSONQO 1233
Db 1165 ydtsktsdvtifagengittkvnkvvrvgldqtkltgtpkltvgnnngkglvidsqngq 1224
QY 1234 NTITGLSNTLANVTNDKGSVRTTEOGNIKDEKTRAASTVDVLSAGFNLOGNGEAVDFV 1293
Db 1225 ntitglslntlanvtnndkgsrvttceqgnlkdedktraasivdvlsagfnlqnggeavdfv 1284
QY 1294 STYDTVNFANGNTTAKVTYDDTSKTSKVYDVNVDDTTIEVKDKKLGIVKTTTLTSTGTG 1353
Db 1285 stydtvnfadgnattakvtyddtsktskvvydvnvddttievdkdklgvktttltstgtg 1344
QY 1354 ANKFALSNQATGDAVKAASDITVAHLNLTSGDIQTAKGASOANNSAGVVDAGNKVIYDST 1413
Db 1345 ankfalsnqatgdalvkasdivahlntlsdgiqtakgasqannsagvyvdagnkviydst 1404
QY 1414 DNKYQAKNDGTVDKTEVAKDKLVAOAQTPDGTFLAQMNVKSVINKQVNDANKKOGINE 1473
Db 1405 dnkyqakndgtvdktevakdklvaqaqtpdgtflaqmnmvksvinkqvnndankkqgine 1464
QY 1474 DNAFVKGLEKAASDNKTKNAAVTVGDLNVAQAQTPLTGAGDTGTTAKKLGETLTIKGGQTD 1533
Db 1465 dnafvkglekaasdnktnaavtvgdlnvavaqtpltgagdtgtttakkllgetltikggqtd 1524
QY 1534 TNKLTNDNIGVAGTDFGVKLANDLNLSVNAAGTPIDEKGFSEFVDANGQAKANPVL 1593
Db 1525 tnkltndnigvagtdfgvklanldnlsvnaggtkidekgsfvsfvdssgqakantpvl 1584
QY 1594 SANGLDLGGKVISNVGKGTDKTDAAVQQLNEVRNLLGL---GNDNADGNOVNADIKKD 1650
Db 1585 sangldlggkvisnvvgkgtktdaavvqqlnevrnllglgnagnadnagqvnliadikkd 1644
QY 1651 PNSGSSNRTVIKAGTVLGGKGNNDTEKLATGGVQGVGVDKNGANGDLSNVVWVKTKDGS 1710
Db 1645 pnsqssnrtvikagtvlggkgnndteklatggvqgvdkdngangdlsnvwwtkdkgs 1704
QY 1711 KALLATYNAAGQTNVYTNPAEATDRINEQIRFFHVNDGNQBPVQGRNGIDSSASGK 1770
Db 1705 kallatynaagqtnylnnpaeatdrineqirffhvdngnqbpvqgrngidssasgk 1764
QY 1771 HSAVIGFQAKADGEAAVAGRTQAGNOSIAIGDNAQTGQOSTAIGTGNVVVAKHSGAI 1830
Db 1765 hsaavgfqaadgeaavaigrtqagngsiaigdnaqtgqgsiaigtgnvvvaghsgai 1824
QY 1831 GDPSTVKADNSVSGNNOFTDATQTDVFGVCNNITVTPESSVALGNSAISAGTHAGTQ 1890
Db 1825 gdpstvkdnsysvgnnoftdatqtdvfgvgnntlvcesnvalgsnsaisagthagtq 1884
QY 1891 AKKSDGTAGTCTTTAGATGTVKFGAQTAVGAVSFGASGAERRIQNVAAGEVSATSTDAVN 1950
Db 1885 akksdgtagtttagatgatkvgfaggtavgvavsgasgaerriqnvaagevsatstdavn 1944
QY 1951 GSOLYKATOSIANANWELDRHTQHNENKANAGISSAMAMASMPQAYIPGRSMVTGGIATH 2010
Db 1945 gsolykatgsianatneidhrihqnenkanagissamampqayipgrsmvtggialth 2004

QY 2011 NQOGAVAVGLSKLSDNGOWFKINGSADTQGHVGAAGVAGPHF 2053
Db 2005 ngqgavavglsklsdngqvwfkingsadtqghvgaavgagthf 2047

RESULT 3

AAB69133
ID AAB69133 standard; Protein; 1992 AA.

XX AAB69133;

XX 24-APR-2001 (first entry)

XX M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

DE Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59100, AAF59101.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis

XX Example 3; Fig 2A-W; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)
XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
XX The 200 kDa outer membrane protein (II) has antibacterial activity and
XX can be used in vaccines. (II), and its truncated versions, are used as
XX immunogenic compositions and vaccines to protect against M. catarrhalis
XX infections, particularly otitis media in humans. (II) is also used as
XX antigen in immunoassays for detecting specific antibodies (Ab), and to
XX generate Ab. (I) are used for recombinant production of (II) and its
XX fragments are used as probes for identifying/cloning 200 kDa protein
XX genes from other strains, and for diagnostic detection of M. catarrhalis.
XX (I) makes possible production of large amount of recombinant immunogens.
XX Expression of truncated versions of (II) reduces toxicity of the protein
XX towards the Escherichia coli host. The present sequence represents the
XX M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is
XX used in the exemplification of the present invention.

XX Sequence 1992 AA;

Query Match 71.5%; Score 7408; DB 22; Length 1992;

Best Local Similarity 74.2%; Pred. No. 0;

Matches 1550; Conservative 134; Mismatches 218; Indels 186; Gaps 28;

QY 56 VIGATLNGSAVAQOITTK-TEIGTNNKINNTLKGDALATGEASTAFSLSKAQSQAIAI 114

Db 1 viga1sgsayaqkdkthaiqeqprts--gtakadgdraialgenanaggggaial 58

QY 115 GSVAPDPNNNGSNVGVSHAKGNESIAIGDVLAEAGDASIAIGSDDLXL-----PKN 165

Db 59 gsnaktvngssaldkigtatgqesialggdvkasgdasialgdlldqhgngphkg 118

QY 166 LDLNKNEFKLIHGHEILKKIOTSTDGKIKYRRTAQGHASTAVGAMSVYAQGHFNSAFGTY 225

Thu Sep 13 14:19:18 2001

```
XX PN W09634960-A1.
XX PD 07-NOV-1996.
XX XX 29-APR-1996; 96WO-CA00264.
XX PF 26-MAR-1996; 96US-0621944.
XX PR 01-MAY-1995; 95US-0431718.
XX PR 07-JUN-1995; 95US-0478370.
XX XX (CONN-) CONNAUGHT LAB LTD.
XX PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX DR WPI; 1996-506162/50.
XX DR N-PSDB; AAT38740.
XX XX Moraxella outer membrane protein - useful as immunogen in protective
XX PT vaccine and for diagnosis
XX PS Claim 14; Fig 6; 109pp; English.
XX CC An approx. 200 kDa outer membrane protein (AAW04505) can be
XX CC isolated from Moraxella catarrhalis otitis media strain 4223
XX CC by electroelution, or expressed from a gene (see also AAT38740)
XX CC obtd. from a strain 4223 genomic library. Natural or recombinant
XX CC outer membrane protein is useful as an immunogen to protect
XX CC against infection by Moraxella, esp. M. catarrhalis. It can
XX CC also be used to detect antibodies, esp. for differential diagnosis
XX CC between bacteria that cause similar symptoms, and also useful as
XX CC a carrier for other antigens and used to raise antitumour
XX CC antibodies for conjugation to therapeutic agents.
XX SQ Sequence 1992 AA;

Query Match 71.5%; Score 7405; DB 17; Length 1992;
Best Local Similarity 74.2%; Pred. No. 0;
Matches 1549; Conservative 135; Mismatches 218; Indels 186; Gaps 28;

Qy 56 VIGATLNGSAVAQITTK-TEIGTFNKINNTLKGDALATGEASTAFGSLSKAOGSOAIAI 114
Db 1 migtatlsayaqkdtkhiaiegnqprrs--gtakadgdraiaigenanaagqgaia 58
Qy 115 GSVKPDPPNNGSNGVSHAKGNESIAIGGDVIAEGCDASIAIGSDDLYL-----PKN 165
Db 59 gssnktvngssldkigtatgesialggdvkasgdasialgsdhlldqghnphkphg 118
Qy 166 LDLNKEFHKLHIGHLKKTSTDGKIKYRTRAQGHASTAVGAMSAYAOGHFSNAGFTY 225
Db 119 -tclnd---linghavlkkeirsskndvkyrrttasghas cavgamsyagghfsna fgr 174
Qy 226 ATAAEAYSILAVGLAAQATKQSSIAVGSNAKANAFAATAIGGNTVNLGRVALFGFSQIL 285
Db 175 alaksayslavglataaegqstiaigsdatsslgalaigagtraqlggsialggsvvt 234
Qy 286 DRDNDTASAVYPLGKTLADQKATQCGDSTDFISGNNSNNNSIRKLIINVCAGSRDT 345
Db 235 qsdnns-rpaytpntqaldpkfqatnntkagpl-sig-----snsikrkiinvagvntk 287
Qy 346 DAVNVAOLKLVLELA-NRKITFGGDGNNNSNVERGLGNTLTIKGDATQNALTEANIGV 404
Db 288 davnvaqlaavvkwakerritfq--gddnstdvkigldntlitkggaetnaltddnngv 345
Qy 405 --TDGNGLKVLAKELTGITSV-----SATNKITVNTNNNNAELOSGGLTFS-PIITGK 456
Db 346 keadnsglkvkiaktlnnltevtnttlnattvkvgssssttaeillsdltftqntgsq 405
Qy 457 -TDKTVYSIDGLKFTNDNSNIATKGTTRITKKIKGACFNDCGVDSKPKYLDNEKLVGNS 515
Db 406 staktyvgvngvkftnn----- 422
```



```
Db 1465 gggtdntkltannlgnvagtgtvkvklakdltnlnsvnagctklddkgvsfvdsggaka 1524
Qy 1589 NTPVLSANGLDLGGKVISNVGKGTDTDAANVQQLNEVRLGLL---GNDNADGNQVNI 1645
Db 1525 ntpvlsangldggkvisnvvgkgtktdaanvqqlnevrnlilglnagndnagngvnia 1584
Qy 1646 DIKDPNSGSSNRVTKAGTVLGGKGNNDTEKLTATGCVGVQVDPKDNAGNGLSNVWKT 1705
Db 1585 dikkdpnsgssnrtvlikagtvlggkgnndtektlatggigvgvdkdngangdlsvvwt 1644
Qy 1706 QKDGSKALLATYNAAGTNTVNNPABIDRINEOGIRFFHVNDGNOEPVVOGRNGIDS 1765
Db 1645 qkdgskallatynaagntvltmpaealdrineeqgrrffhvdngnqepvvvgngids 1704
Qy 1766 SASGKHSVAIGFOAKADGEAAVAIGROTQAGNQSIATIGDNAQATGDOSIATIGTNVAGK 1825
Db 1705 sasgkhsvaigfakadgeaavaigrtqagngsialgdnagatgqdsialgtgnvvagk 1764
Qy 1826 HSGAIGDPSTVKADNSVSGVNNQFTDATQTDVFCVGNNTVTSNSVALGSNAISAGT 1885
Db 1765 hsgaigdpstvkdnsysvgnnqftdatqtdvfgvgnntvtsnsvalgsnaisagt 1824
Qy 1886 HAGTQAKSDCTACTTTTAGATGVKGFAGOTAVGAVSVGASGAEFRRTQNVAAGEVSATS 1945
Db 1825 hagtcakksdctacttttagatgtvkgfagtcavgavsvgasgaeerrlqnvaaagevsats 1884
Qy 1946 TDVNGSOLYKATOSIANATNELDHRHQENKANAGISSAMAMSPQAYIPGRSMVTG 2005
Db 1885 tdvngsqlykatqsiatanatneldhrilhqenkanagissamamspqayipgrsmvtg 1944
Qy 2006 GIATHNGOGAVAGLSKLSNDGQWFKINGSADTQGHVGAAGAGFHF 2053
Db 1945 giathnggavagvlsklsdngqwfkingadtqghvgaavagagfhf 1992

RESULT 5
AAB69137
ID AAB69137 standard; Protein; 1992 AA.
XX
AC AAB69137;
XX
XX 24-APR-2001 (first entry)
XX
DE M. catarrhalis M56 200kDa protein in pKs348 SEQ ID NO:13.
XX
KW Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
XX WO200107619-A1.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-CA00870.
XX
XX 27-JUL-1999; 99US-0361619.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX WPI; 2001-159722/16.
XX DR N-PSDB; AAF59106.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
XX
XX Claim 1; Fig 8A-V; 247pp; English.
XX
XX The present invention describes an isolated and purified nucleic acid (I)
```

```
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis M56 200kDa protein in pKs348, which is given in the
CC exemplification of the present invention.
XX
SQ Sequence 1992 AA;
```

```
Query Match 71.5%; Score 7405; DB 22; Length 1992;
Best Local Similarity 74.2%; Pred. No. 0;
Matches 1549; Conservative 135; Mismatches 218; Indels 186; Gaps 28;
```

```
Qy 56 VIGATLNGSAYAAQOITTK-IEIGOTNKINNTLKGALATGEASIAFGSLSKAQSQAIAI 114
Db 1 migatlsysayaqkdtkthiaigeqnprrs--gtakdgdrailaigenanaggggatai 58
Qy 115 GSVKPDPPNNGSNGVNGSHAKNESIAIGDVLAEAGDASIAIGSDDLYL-----PRN 165
Db 59 gsnkntvngsidkigtatgesiaigdvkasdiaigsddhildqhgphkphkg 118
Qy 166 LDLNKFEHLHHEILKKTOTDGTGKIYRRTTRAQGHASTAVGAMSYAQGHFSNAGTY 225
Db 119 -tlind---linghavlkkeirsskdndvkrrttasghastavgamsyaqghfsnagtr 174
Qy 236 ATAEAAVSLAVGLAAQATKQSSIAVGSNAKAFATTAIGNTVNLGRVALGFSQTL 285
Db 175 ataksaysiaivglaaatgaegstiaigsatsslgailagtrqqlggsalggsvvt 234
Qy 286 DRDNNTDASAVPLGKTLADQYKATROGSDTDFISGNSNNNNNSIRRKIIIVGAGSRDT 345
Db 235 qsdnns-rpaytpntqaldpqtatntkagpl-sig-----snsikrliivgagvntk 287
Qy 346 DAVNVAQLKLVLELA-NRKITFKGDDNNSNSVERGLNTLTIKGDAQTNAITEANIGVV 404
Db 288 davnvaqlavkvkewakeritfq--gddnstdvkigldntlikggaeatnaldnngvr 345
Qy 405 --TDGNGLVKVLAKELTGLTSV-----SATNKITVNTNNNAELQSGGLTFS-PITGTF 456
Db 346 keadnsglkvklaklennltvnttlnattctvkvgsssttaeillsdltftqptgsq 405
Qy 457 -TDKTVYSIDGLKFTNDSNSIATKGTTRITRKKKIGFAGTNDGVDESKPYLDNEKLKVGNS 515
Db 406 stsktvvgvngvkftnn-----
Qy 516 TLNSSLTVNNTTGNKQIQVGANGIKFATVANNVANTSATVGTARITEEIKGFAGTNDGV 575
Db 423 -----aettaaigttritrdrkigfardgvd 448
Qy 576 DEQAPYLDKERLKYGRVEITTDGGINAGNHKITGTGTNGTANTDAVTIKOLDAKPTLNAAG 635
Db 449 ekqpypldkqklkvgsvaitdngidagnkklslnlakssandavtielekkaaktlnag 508
Qy 636 DGISINSNNGDLVDSSGNITPTTYNISVKTTLKNSNGTSGNNKFSVSNNAHDNNSLVATKD 695
Db 509 agisvtpteisvdaksgnvtaptynigvkttelnsdglts--dkfsvksgsgttnslvtaeh 566
Qy 696 LADYLNKNEVTSADSLPFSKVGQNDNSN-NAITVGKDT--NOKTFTNLKLGKNGVNI 752
Db 567 lasylnevrntadsalqsfvkeedddanaaitvakdtknagavsiilkkgkngltvat 626
Qy 753 NRATGTVTFIGDQSNGLTTPKLTVGSD-----TNGN----- 783
Db 627 kk-dgtvtfisqsgsltigkstlndngltvktdneqigvgangikfctnvgnsnpgtgia 685
```


Best Local Similarity 64.1%, Pred. No. 0;
Matches 1553; Conservative 140; Mismatches 250; Indels 481; Gaps 38;

QY 1 MNHIYKVFINKATCFMNAVAEYAKSHSTGGGSCATGOVGSVRTLSFARIALAALVLAVIGAT 60
 | | | | | | | | | | | | | | | | | | | | : | : | : | : | : | : |
Db 1 mnhiykvifnkatcfmnavaeacakahaggsssttagvqgsspvlrltvalailavigat 60

QY 61 LNSAYAAQTTKTKEIOTNKNINNTLKGDALATGEASTAFGLSKAQSOQAIGSVKPD 120
 | | | | | | : | | : | : | : | : | : | : | : | : | : | : |
Db 61 lngsayagn-nsktaftgtgnndn-----sasneasiaigsalahangalaiggskpd 114

QY 121 P-NNGSNGNWGHAKGNESTAIGGDVLAEGDASTAGSDDLYLKPNDLKKNEF----HK 174
 | | : | | | | | | | | | | | | | | | | | | | | : | : | : |
Db 115 prqaanagkaghakgesiaiggdvlaeodaslaigsdldylrdrn-stnskyngllist 173

QY 175 LIHGHEILLKIQTSTDCKIKYRRTPAOCHASTAVGANYSYAOGHFSNAFGTTATATAEAYS 234
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 174 liqhvtlvrqirds-ngsqkyrrtaaeghastavgamayakhafanafigtstaegnysl 232

QY 235 AVGLAAQATKOSSIAVGSNKANAPATAIGGTVTNWLGRCVALGFSGQILDRONNTDAS 294
 | | | | | : | | : | | | : | : | : | : | : | : | : | : | : |
Db 233 avgitakaegytiiaignaqainygalaigadctrdvdiyalgygsqilnnnnnonk 292

QY 295 AYVPLGR-TLADQYKATRQGDSTDIFSIGSNNNNSSIRRKLIINVGSRDTDAVN AOL 353
 | | | | : | | | : | | | : | : | : | : | : | : | : | : |
Db 293 ayvpengsnikskskatcng-----lfaisg-----sstkrklinvgagedtdavnvaql 343

QY 354 KLVBELANRKITTFPGDDNNSNVERG LGNTLTIK-GDAQTNALTE-ANICVVTDGN-GL 410
 | | | | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 344 kavenlakrgittfk-gddngtgvvkkkgeltikgggetqadktidnmniqvvdntnlgl 401

QY 411 KVKLAKELTGTSVS-----ATNKITYSNTNNNAELQSGLTSPITGTDKXTVI SID 465
 | | | | | : | | : | | : | : | : | : | : | : | : | : | : |
Db 402 kvklaklnslgtstvtnltaseaktv-gsgnptaelqsgglttpttinastdtktyvgtd 460

QY 466 GLKFTNDSNSTAKGTRIRTKKKGIPAGTNDGVDEBSPYLDNEKLKVGNSTLNGSLTVN 525
 | | | | : | : | : | | | | | : | : | : | : | : | : | : |
Db 461 glkfcdnsnt-aledttrickdigfkanagtvdenkpyldkhlkvgnstlngngltnv 519

QY 526 NWTG--NQIOVGANGIKFAIVANNVA NTATVCTARTTBEEKIGFAGTNCVDGQAPYLD 583
 | | | | : | | | : | | | : | : | : | : | : | : | : | : |
Db 520 ntligssnkqlqvgaadgklfadnvvnvsen-aakfgttriteeeiffadagdkvkkspslyd 578

QY 584 KERUKVRVEITTDGGINAGNHKTTLGTLINGTANTDAVTIKOLKAOKPTLNAGDGISINSN 643
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 579 kkqlqvvgvitkdsginagdqklsnvkdatdtdtavtykqlk----- 621

QY 644 NGDLVDSGNITTPTYNISVKTTLKLSNGTSGNKNKFVSNAHDNNSLYTAOKDLADYLNKY 703
 | | | | ----- 623

QY 704 NETADSALPSFKVONGDNSNAITVGK-DTNG---KTFNTLLKXGENGVNLITNRATCTV 759
 | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 624 qqdagdalqsfisr-dekqqefitslnyngntpnftetifgegensinsidiakgv 681

QY 760 TFGIDQSNGLTTPKLTVGSPTNG-NRLVIEQVPASDAGNSTKNIKIGLSPTLPFIASPSG- 817
 | | | | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 682 kvgidpinglttpkltvgsdkdgktqviegv--asgndtkniirgisptlpisitnaagv 739

QY 818 RNTALGWTI-EEKDKSNAASIDDVLNAGFNLUKNNGKDOKDFVSYTDVDFIDGNATTATVT 876
 | | | | : | | | : | | | : | | | : | : | : | : | : | : | : |
Db 740 rtteeqnrtisedcksaaslgdilntgfnlknnsvsfystyntvdfidgnattaktvt 799

QY 877 YDEANOPSKVAYDVNVPEKTELINGDNCK-KOLCVKTIKLTETSTNGCATMTFFSTDDHAL 935
 | | | | | | | | | | | | | | | : | | : | : | : | : | : | : |
Db 800 ydcnqtskvtydvnnvedktiellcgdnkgknigvkttltttbaangkatsfdmdal 859

QY 936 VKASDIAGNLNTLAEIHHTKGTANTALQOTPVKRVENDKRADDONTATTVGKDGST-SGK- 993
 | | | | | | | | | | | | | | : | : | : | : | : | : | : |
Db 860 vnakdieenintlakeihttcgtaadtalqtffkvk-----dgatdetitvkgdgtqngkt 915

QY 994 VNTLUKKGKNGLDIKTDKGTVTFTPIGINTQSGLKAGDSTTLNNGLSIKNTASNBOIQVGA 1053
 | | | | | | : | : | : | : | : | : | : | : | : | : | : | : |

Qy	1	MNHIIYKVFINKATGTEPMAYAEYAKSHSTGGSCATQGVQSVRTLSPFARIALA-----	53
Db	1	mnkfnvwnvmtqvvvseltrthtk--rasatvevatilalfatvganatedeel	58
Qy	54	-----VLVIGA-----	59
Db	59	dpvvrtapvlsfshsgektgekvtensnwglyfdnkgvilkagaltikagdnlikgntd	118
Qy	60	-TLNGSAYAQOI-----TKIEIG-----OTNKI	82
Db	119	estnasftyyslkkdlitstsvateklsgangkdvltisdanglkiaktgngnvhingl	178
Qy	83	NNTLKGDALATGASTAFSGSLSKAQGSQAIACTGVKPPDPNNGSNGVCSHAKGN-ESIAI	141
Db	179	dstlpdvtntgvlssssftpndevktra--atvkdvlnagwnlkgaktgngvsvdl	235
Qy	142	GGDVLAEGDASIAIGSDDL-----YLPKNLDLKNFPHKLIHGHEILKK	184
Db	236	---vsaynnvefitgdntldvltakengkttevkftpktsvikekdgklftgke---	288
Qy	185	IQTSTDGKIYRRTRAGHASTAVGAMSYAQHFSNAFTYATAPAAAYSLAVGLAAQATK	244
Db	289	-nndtnkvt-----sntatdntdegng-----lvtakavid-avnkagwrvrk	328
Qy	245	QSSSTAVGSNAKANAFATAJGGNTVNLGRGVALGFGSILDRDNNTOA-----SAYVPL	299
Db	329	ttt-----angngofatvasgntvntfiesgdgttas-----vktlungngiltkydakvgd	379
Qy	300	G-----KTLADQYKATROGSDTDFISIGNSNNNSSIRKLIINVAGSRDRTDAVNAQ	352
Db	380	gikfadskkivadtaltvtvgk--vaeiakedd-----kklvnag-----	419
Qy	353	LKLVEELANRKITPKGDGDN-----SNSVERGLGNTLTIKGDAQTNALTAEANIGVVTD	406
Db	420	-dlvtalgnlswkakaeadtdgaleglskdqevkagetvftkag-----knlkvkd	470
Qy	407	GNGLKVKLAKELTGLTTSVSATNKITVTSNTNNNAELQSGGLTFSPITGKTOKTVYSIDG	466
Db	471	ganfyslqdalgtls-----ltlgttgn-----ggnadaktvinkdg	508
Qy	467	LKFTNDSNSIATGTRTRIKKKIGFAGTNDGVDESKPYL---DNEKLKVGNSFT---LN---	518
Db	509	lritpagn-gtfgtntisvckdglkagnkaftnvagslrayddanfvlmnsatdlnrh	567
Qy	519	-----SGSLTVNNTTGNKIQIOVGANGIKFATVANNVANTSATYGTARITEEIKGF-AGTN	572
Db	568	vedaykgllnlnekanakplv-----tdstaatvgdlr-----klgwvsvtk	610
Qy	573	DGVDEQAPVLDK--BRLLKVRVEITTDGGINAGNHKITGLTNGTANT-----DAVTI	622
Db	611	nglkeesnqvkqadevlftgagaatvtsksengkhti---tvsvaeckadcglekdgdti	667
Qy	623	KQLKDAKPTNLNAGDGISINSNNGLDVDSGNIITPTTVNISVKTTLKNSNTGSGNNKFSVS	682
Db	668	klkvdnqnt-----dnvltvgngntavtkgget-----vktgatdad-----rgkvtvk	712
Qy	683	NAHDNNS-----LVTKADLADYLLKNVNETADSLPEKVKONGDNS---NNAITVGKDTNGKT	736
Db	713	datandadkkvatkvdaaln-----saatfvkntentltsidednpldngkddalka	765
Qy	737	FNTLKLK-GEN-----GWNIT-----TNRATGTVTFGIDQSGNGLT---TPKLTVG	777
Db	766	gdtltfkagknklkvrdgkhnifldaknlevktakvsdtltlgnntpggttaltapkvnit	825
Qy	778	SDTNGNRLVJEQVPSADGNSGTNNI--IKGLSPTL--PSTASPSGRNIALGNTIEBKDSNA	834
Db	826	staadgnfake-----tadasgknvylkglatlitlepsagakssh-----vdlmvdattkksna	879
Qy	835	ASIDVDVLNAGFNLLKNGKDKDFVSTVDYDFIDGNATATTVYDEANQTSKVAYDVNVNDE	894
Db	880	asiedvllragwnlqgnngnnvdyvatydtvnfdldstgtltvltq--kadgkgadvkiga	937
Qy	895	KTIELTDGNGKKQLGVKTKTLTETSTNGNATTFSTFDDDH-----ALVKASDTF-----	941

1790	GRO	TAGN	QSG	TAIG	DN	QAQ	TG	DQ	S	T	A	I	G	T	G	N	V	A	K	H	S	G	A	I	G	D	P	S	T	V	K	A	D	N	S	Y	S	V	G	N	N	Q	1849																		
2051	g	r	q	t	a	g	n	q	s	i	a	i	g	n	q	s	i	a	i	g	n	v	t	g	k	h	s	a	i	g	d	p	s	t	v	k	a	d	n	s	y	s	v	g	n	n	q	2110													
1850	F	T	D	A	T	O	T	O	T	D	V	F	G	N	N	I	T	V	T	E	S	N	A	L	G	S	N	S	A	I	S	A	G	T	H	A	G	T	O	A	K	S	D	G	T	A	G	T	T	T	1909										
2111	f	i	d	a	t	q	d	v	f	g	y	n	n	i	t	v	t	e	s	n	a	l	g	s	a	i	s	a	i	s	a	g	t	h	a	g	t	q	k	s	d	g	t	a	g	t	t	t	2170												
1910	V	K	G	F	A	G	O	T	A	V	G	A	V	S	Y	G	A	S	G	A	E	R	R	I	Q	N	V	A	A	G	E	V	S	A	T	S	T	D	A	V	N	G	S	O	L	Y	K	A	T	Q	S	I	A	N	A	T	N	E	L	D	1969
2171	v	k	g	f	a	g	t	a	v	g	a	v	s	y	g	a	s	g	a	e	r	r	i	q	n	v	a	a	g	e	v	s	a	t	s	t	d	a	v	n	g	s	q	l	y	k	a	t	g	i	a	n	a	t	n	e	l	d	2230		
1970	H	R	I	Q	H	E	N	K	A	G	I	S	S	A	M	A	S	M	P	O	A	Y	T	P	G	R	S	M	T	V	G	G	I	A	T	I	N	G	O	G	A	V	A	G	L	S	K	L	S	D	N	G	Q	W	2029						
2231	h	r	i	q	h	e	n	k	a	g	i	s	s	a	m	a	s	m	p	o	a	y	t	p	g	r	s	m	t	v	g	g	i	a	t	i	n	g	o	g	a	v	a	g	l	s	k	l	s	d	n	g	q	w	2290						
2030	V	F	K	I	N	G	S	A	D	T	O	G	H	V	G	A	N	A	V	G	A	G	P	H	F	2053																																			
2291	v	f	k	i	n	g	s	a	d	t	o	g	h	v	g	a	n	a	v	g	a	g	p	h	f	2314																																			

RESULT

7

AAR99393

AAR99393 standard; Protein; 2353 AA.

AC

AAR99393;

15-JAN-1997

(first entry)

DE

Haemophilus adhesion protein HA2.

XX

Haemophilus adhesion protein; HA2; hsf protein; vaccine.

XX

Haemophilus influenzae type b strain C54.

XX

W09630519-A1.

XX

03-OCT-1996.

XX

22-MAR-1996; 96WO-US04031.

XX

24-MAR-1995; 95US-0409995.

XX

(UYSL-) UNIV ST LOUIS.

XX

(UNIW) UNIV WASHINGTON.

XX

Barentkamp SJ, St Geme JW;

XX

WPI; 1996-455364/45.

XX

N-PSDB; AAT41476.

XX

Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in vaccines against H. influenzae infection.

XX

Claim 5; Page 66-73; 120pp; English.

XX

Haemophilus adhesion protein HA2 (AAR99393) is associated with the formation of surface fibrils involved in adhesion to various host cells; it is also referred to hsf (Haemophilus surface fibrils).

XX

Its amino acid sequence was deduced from a genomic DNA clone (AAT41476) derived from Haemophilus influenzae type b strain C65.

XX

Large quantities of recombinant HA2 can be produced in transformed prokaryotic or eukaryotic host cells, for use in vaccines against H. influenzae infection.

XX

Sequence 2353 AA;

XX

SO

Query Match	11.6%;	Score 1206;	DB 17;	Length 2353;
Best Local Similarity	23.8%;	Pred. No. 4.2e-45;		
Matches	627;	Conservative 295;	Mismatches 853;	Indels 856;
				Gaps 114;

Db 938 klsvikdhngk1ftg-kolk-----danggatvseddgdktgtglvtaktvidavnsksw 991
Qy 942 -----AGNLTAE--EIHTRGTANTALQTFTVKKVDEN----- 974
Db 992 rvtgegataetgavnaetaetvtsvfnkgntatta-----tvskdngninvkydvn 1047
Qy 975 -----DK--ADDTNATVGVKDGTSKGV-----NTLKLKKNGL----- 1005
Db 1048 vgdglkigdkkivadttltv-----tgkvsvpagansvnnkklvnaeglatalnlnls 1103
Qy 1006 -DIKTDKGTFTFGINTOSGLKAGDSTTLN--NGLSIKN-----TASNEQIQVGADGVKE 1058
Db 1104 wtakadkyadesegedqevkagdkvfkagknkfkqskdfyslqdtlgtltsltl 1163
Qy 1059 AMVNVGVGAGIDGTRITRDEIGFTGTNGS---LDKSKPH---LSKDCINAGGKKTINI 1112
Db 1164 ggtangrndtg---tvinkdgltitlangaaagtdaangntisvtkdgisagnkeilnv 1219
Qy 1113 QSG-----EIAKNSHDAVTGGKYIDLKTELENKISSTAKTAQNSL 1152
Db 1220 ksalktykdtntadqtdqkfehaavkanevefvgk-----ngatvtsaktcdngk 1270
Qy 1153 H-----EFSVADEQGNFTVSNPYSSVDTSK-----TSDVI 1183
Db 1271 hvtldvaeakvgdglektdgkiklvndtgnlltvdatkgaasvakgefnavttat 1330
Qy 1184 TFAGENG-----ITTKVKNKGVVRVIGIDTKGLT----- 1211
Db 1331 taqgtnanerkvvvkgsgngatataetdkkvatvgdvakaindaatfvkvvendseatldd 1390
Qy 1212 -----TPKLTGVGNNG 1222
Db 1391 sptddgandaikagdtltlkagknkvrkgdnitfalandsvksatvsdklsigtgn 1450
Qy 1223 KGIVINSQNGN-----TINGLSNLTANTNDKSGVTTTEQGNIIKDEDTR 1269
Db 1451 kvnlcsdckglnfakdsktgdanlhngiaslttdtllnsgat-tnlgngnigtnekr 1509
Qy 1270 AASTVDLVSAGFNLOG-----NGEAVDFVSTVDTVFNFANGNTTTAKVTYD--DTSKT 1319
Db 1510 aasvkdvlmagvrvvgkpasannqvenidfvyatdytdvfvgdkdtsvtveskdngkr 1569
Qy 1320 SKVYDVNVDDTTEVKDKKLGKVTMTTSTG-----TGAN-KFALSNOAT-----G 1365
Db 1570 tev-----kigaktsvikdhngklftgkelkdannngvtvtetdgkdeg 1613
Qy 1366 DALVKASDIVAHLNLTLSGDIOTAKGASOANSAGTYD-----ADGNKVIYDSTDNKYY 1418
Db 1614 nglvtakavidavnkagvrvtkt-gangqnddfatvasgtnvtfadngnttaevt----- 1667
Qy 1419 QAKNDG--TVDKTEVAK-----DKLVAQA---QTPDGTILAQMVKSVINKEQVNDANK 1467
Db 1668 -kandgsitvkynkvvadgklldgdklvadtvtltvaagkvtaqn-----ngdgk 1716
Qy 1468 KOGINEDNAFVKLEKAASDNKNTNAAVTVGDLNNAVOTPLTFAGDTGTTAKKLGTELTI 1527
Db 1717 k-----fvdasgladalnklswtatagkegtgevdpansaqg-----evkagdkvtf 1763
Qy 1528 KGGQTDTKLTDNNIGVVAGTGDGFTVKLAKDLTNLSV-----NAG-----GTTKIDEGIS 1578
Db 1764 kag-----dnlkikqsgkdfyslkkelkdltsvefkdangngtgsstklkdkdalt 1814
Qy 1579 FVDANGQA-----KANTPVLSSANGLDLGGKVISNKGKTK-----DTPDAAN 1619
Db 1815 itpangagaagantantitsvtdkgdisagnkavtnvsvglkfgdghtlangtvadfehy 1874
Qy 1620 VOQLNEVRNLLGLGNDN-----ADGNQVNIAD-----IKKDPNNGS-----SSN 1658
Db 1875 dnaykdllnldekagdnptvadntaatvdlrglwwisadkttgepnqeynaqvnrnan 1934
Qy 1659 RTVTKAGT-----VLGG-----KGNNDTEKLATGG-----VQVG----- 1687

Db 1935 evkfksnginvsqtklntgrvitfelaakgevvksneftvknadsetnlvkvqgdmysk 1994
Qy 1688 VDKDGNAGDLSNVVWVKTK-----DGSKKALIATYNAAGQTNVYNNPAAEID 1736
Db 1995 edidpatskpmtg---ktekykvengkvvsangsktevtltnkgsq---yvtn-qvad 2046
Qy 1737 RINEGIRIFFHVNDGNQEPVVOGRNGIDSSASGKHSVAIGFOAK-----ADGEA 1785
Db 2047 alaksqfel-----gladaaeakafaesakdkqlskdkaetvnaahkv 2090
Qy 1786 AVAIGROTQAGNQSTIAIGDNAQATGDO-----SIAITGTVNVAGKHSIGAIDPSTVK 1837
Db 2091 rfanglnlcnkv---saatvestdangdkvttfvtkdvelplqlty---ntdangnkivkk 2144
Qy 1838 ADNSYSVGNNNQFTDATQTDVFGVGNNTITTESNSVALCSNSAISAGTHAGTQAOKSDGT 1897
Db 2145 adgkw-----yelnadgtasnkevtlgnvdangkk-----vvkvteq 2182
Qy 1898 AGTTTATAGATGVKGFAGQTAVGAVSVG-----ASGAERRIQNVAAGEVSATS 1945
Db 2183 adkwyttnadgaadkktkgevsndkvstdekhhvridpnngsqngkvvidnvangeisats 2242
Qy 1946 TDAVNGSOLY---KATQSTANATNELDHRILHONENKANAGISSAMAMASMPQAYIPGRSM 2002
Db 2243 tdaingqlayavakgvtlnlagvnnlegkvkvkradagtasalaasqlpdatmpgksm 2302
Qy 2003 VTGGIATHNGQCAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGVAGPHF 2053
Db 2303 valaegssyvgqnglaigvrisdngkviirlsgttngskgtvaagvgvq 2353
RESULT 8
AAB23860
ID AAB23860 standard; Protein; 2411 AA.
XX AAB23860;
AC AAB23860;
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
KW non-typeable Haemophilus influenzae; antinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
PW WO2000055191-A2.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-618897/59.
DR N-PSDB; AAA92499.
XX
XX Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Haemophilus influenzae
PT infection
XX
PS Claim 1; Fig 24; 275pp; English.
XX
CC The present sequence represents a Haemophilus influenzae adhesin
CC (Hia) protein from the type c Haemophilus influenzae strain API.
CC Hia genes and proteins have antinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An

CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.

XX Sequence 2411 AA:

Query Match 11.6%; Score 1205; DB 21; Length 2411;
Best Local Similarity 23.8%; Pred. No. 4.8e-45;
Matches 630; Conservative 299; Mismatches 863; Indels 880; Gaps 115;

QY 1 MNHIYKVFINKATGTFMAVAEYAKSHSTGGSCATQGVSVRYLSPTAALAA-----53
DB 1 mknifnvwmtqtvvvseltrthck--rasatvetavlatllfatvganatedeel 58
QY 54 -----VLVIGA-----TLN 62
DB 59 dpvvrtpvlsfhskdtegekevtensnwgiyfhngkvlgkagaitlkagdnlikqstn 118
QY 63 GSAYAAQI-----TPIKIEIG-----QTNKNNTL 86
DB 119 assfysikkldltsvateklsfngangkvditsdangliklaktngnvhngldstl 178
QY 87 KGDALATGEASTFGSLKAGCSQAIAIGSVKPPNNGSNGVSGSHAKGN-ESTIAIGDV 145
DB 179 pdavntvgvslssstfpndvektra--atvkdlvngwnlkagktaggnvesvdl---v 232
QY 146 LAEGDASIAIGSDDL-----YLPKNLDLNEPHKLIHGHIEILKKTQTS 188
DB 233 saynnvefitgdknldlvltakengkttvfktpktsvikekdgkiftgke-----nn 286
QY 189 TDGKTKYRTRAGHASTAVGAMSAQAQHFNSAFCGYATAEAAYSLAVGLAAQATKOSSI 248
DB 287 dlnkvt-----sntatdntdegng-----lvakavid-avnkagrvkttt- 327
QY 249 AVGSNAKANAFAATAIGGTVVNLGRVALGFGSOILDRDNTDA-----SAYVPLG--- 300
DB 328 ---angngdfatvasgntvtfesgdgtas-----vktngngitvkdakvgdglkf 379
QY 301 ----KTLADQYKATROGSDTDIFSIGNSNNNSSIRRKLIINVGAGSRDADVNVAAQLKV 356
DB 380 dsdkkivadtaltvtggk--vaeiakedd-----kkklvngag-----dlv 418
QY 357 EELANRKITFKGGDNNNSNVERGLNTLTIKGAQTNALTEANIGVYVTDGNGLVKVIKAK 416
DB 419 talgnlswkakaeadtdtdgalegiskdqvketvtfkagknkvkdganfysiqd 478
QY 417 ELTGTLTSVATNKITVSTNTNNNAELQSGGLTFSPITGKTOKTVYSIDGLKFTNDSNSI 476
DB 479 altglts-----itlgtttn-----ggndaktvinkdglittpagng- 515
QY 477 ATKGTTRITKKIKGFAGTNDGVDESKPYL---DNEKLKVGNSP-----LN-----SGSLT 523
DB 516 gttgntitsvtkdglkagnkainvasglsrayddanfdvlnnsatdlnrhvedaykglln 575
QY 524 VNNTTGNKIQVQANGIKFATVANNVANTSATVGTARTITEEKIGF-AGTNDGVDEQAPYL 582
DB 576 lneknanqplv-----tdstaetvgdlr-----klgwsvstkngtkeesngv 618
QY 593 DK--ERLKVGRVEITTDGGINAGNHKTTGLTNGIANT-----DAVTIKQLKDAKPTL 632
DB 619 kqadeviftgagaatvtsksengkhti---tvsvaetkadsglekdgdtiklkvdnqnt- 674
QY 633 NAGDGISINSNNGDLVDSNGITPTTYNISVYKTKLNSNGTSGNNKFSVSNIAIDNNS--- 689

DB 675 ---dnvltvgngtavtkggfet-----vktgatdad-----rgkvtvkdandadkk 720
QY 690 LVYAKDLADLYLNKVNKTADSALESPFKVQNGDNS---NNALTVCKDNTNGTFTNTKLK-GE 745
DB 721 vatvkdvatain-----saatfvktenlttsidednptdngkddalkagdtltfkagk 773
QY 746 N-----GVNIT-----TNRATGTVTFIGDOSNGLT--TPKLTVGSDTNGNRLVI 787
DB 774 nlkvkrdgnkntfdlaknlevktakvsdtltiggnptggttatcpkvnitstadnfak 833
QY 788 EQVPSADGNSTKNT-INKLSPTL--PSIASPSGRNALGNTIEEKDSNAASDDVNLNAG 844
DB 834 e---tadasgsknvykgiattltleptasagakssh---vdlinvdatkksnaasiedvlrag 887
QY 845 FNLXNNGKDKDFSVTYDTVDIFDGNATTATVYDEANQTSKVAYDVNVVDKTIETLFGDNG 904
DB 888 wnlgggnnvdyvatydlvlnftddstgtttvtvtq--kadgkgadvkigaktsvikhng 945
QY 905 KQOLGVKTKIKLTETSTNGNATFTSTDDDH-----ALVKSADI-----941
DB 946 klftg-kalk-----danngatvseddgkdtgtglvtaktvidavnskgsrvrtvgeetae 999
QY 942 -----AGNLNTLAE--EIHHTTKGTANTALQTFVAKVDEN-----D 975
DB 1000 tgaatavngnaetvsgtsvfnkgnatta---tvskdngnlnvkydvnvgdglkigd 1055
QY 976 K--ADDTNAITVGKDGTSKGV-----NTLKLKCKNGL-----DIKTKDGT 1014
DB 1056 kklivadltltlv---tgkksvpagansvnnkklvnaeaglatalnlnslwtakadyad 1111
QY 1015 VTFGINTQSLKAGDSTTLN-NNGLSIKN-----TASNEIQVQVADGVKFAFVANGVGA 1068
DB 1112 gesegtdedvkgagokvtfkagknkvkqsekdfyslqdtltgltstlgttngandrnt 1171
QY 1069 GIDGTTIRIRDEIGFTGINGS---LDKSKPH---LSKGINAGGKKIINIOS-----1114
DB 1172 g---tvlnkdgltitlangaaagtdaangntisvktgdigsaenkeifnvksalktykdt 1227
QY 1115 -----GETAK-----NSHDA-----VTGGKIVDLK-----1134
DB 1228 qntagatqpaantaevakqdlvdltkpatgaagngadadakpdtatvtdglglgwisa 1287
QY 1135 -----TELE-----NKISSTAKTAQNSLH-----1153
DB 1288 kktadetqkhefaavkvnanevefvgkngatvsaktddngnkhvtidvaekvgdglekd 1347
QY 1154 -----EFSVADEOGNNFTVSNPYSSYDTSK-----TSDVITFPAGENG-----ITTKVNK 1197
DB 1348 tdkkiklvtdntdgnlnllvtvdatkgsavakgefnavtttdatgaqtnanergkvvkvgsn 1407
QY 1198 GVVRVGIDQTKGLT-----1211
DB 1408 gatateldkkkvatvgdvakalndaatfvkvvenddsatidsptddgandalakagdtl 1467
QY 1212 -----TPKLTGVNNGKGVINSQNGQ-----1234
DB 1468 kagknkvkrdgnkntfalndlsvksatvsdklsigtngknvknitsdtkglnfakdskt 1527
QY 1235 -----TITGLSNTLANVNDKGSVRTTEQGNIIKDEKTRAASIVDVLSAGNLOG---1285
DB 1528 gddanhlnglasltltdllnsgat-cnlgngngitdnekkraasvkdvlvlnagvnvrgvp 1586
QY 1286 -----NGSAVDVSTYDTVNFANGNTTTAKVTYD--DTSKTSKVYVDVNVDDTTIEVKDK 1338
DB 1587 asannqvenlfdvtydtdvfsgdkdtavtveskdngkrtev-----1630
QY 1339 KLGVTQTTTLTSTG---TGAN-KEALSNAQAT-----GDALVKASDIAVHLNLTUSD 1384
DB 1631 kigaktsvikhngklftgkelkdannngvtvtetdgddegnglvtakavidavnkagwr 1690
QY 1385 IQTAKGASQANNSAGYVD-----ADGNKVIYSTDNKYQAQKNDG--TVDKTKVEAK- 1434
DB 1434

558	ngastkktkdgl	ltsangangaatadadk	ikvasagisag	nkavknvsvg	lkkfgdan	717		
1621	QQLNEVRNLLGL	NDNADGNQVNIADIK	KDPNSGSSNRTVI	KAGTVLGGKGNND	EKLA	1680		
718	-----npl	tsasdnit-----	-----kgyd	daykgl	738			
1681	TGGVQGVGDK	-----DCNAN--	WKTQDKGSKALLAT	YNAAGOTNYVTN	NP	1731		
739	tnldékadqk	tltvadntaatv	dgrlgwv	isadkktg	elnkeyna-----	786		
1732	AEALDRINEQGI	RFHHVNDGNQEP	VPVQGRNGIDSS	---ASGKH	SVAIGFOAKADGEA	VA 1788		
787	-----qvr	naevkfsg-ngih	vsfgktvngrr	reit---fel	akdena-----	835		
1789	IGROTQAGNQSI	ATGDNAQTGDSIAI	GTGNVVAGKHS	GAIGDPSTV--	KADNSYSVGN	1846		
826	-----ia	fgygskalr	tdntvaigtgn	vvnaeksgaf	gdpnyledkaggs	ya 874		
1847	NNQFTDATOTD	VFVGNN-----	-----ITV--	-----TES	NSVAL	1875		
875	dnrit-sá	ntfvlngv	nakýkangdv	tetvtvkdkg	kettvtvpka	lgatvensvyl 933		
1876	GSNSAISAGTH	AGTQAKSDCAGT	TTTAGATGTGV	KGFAGOTV	AGVSVGASGAE	RRION 1935		
934	gnk-----	statdkgn	ksdtagnttt	agttglvng	fageta	hqvsvgsgsgeerr	iqn 990	
1936	VAAGEVSTATSD	VNGSOLYKATQSI	ANATWELDHR	HTHONENKAN	AGISSAMAS	MPQA 1995		
991	vaageisat	sda	lngsqlyavak	gvtn-----	lagqv	nvkgkradag	casalaasq	lpqa 1046
1996	YTPGRSMVTGG	ITATHNGGAV	AVGLSKLSDNG	QWVFKING	SADTQGHV	GAAGVAG	GFHF 2053	
1047	smpgksmvs	iaegssyqg	gnqgaigv	srisdng	kvilrlsg	ttnsqgk	tgvaagv	gyqw 1104
RESULT 10	AAB23859 standard: Protein: 1104 AA.							

The present sequence represents a Haemophilus influenzae adhesin (K22) protein from the non-typable Haemophilus influenzae (NTHi) strain K22. Hia genes and proteins have anti-inflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher percent of recovery than a full-length protein.

Query Match	9.9%	Score 1021;	DB 21;	Length 1104;
Best Local Similarity	27.5%;	Pred. No. 2.3e-307;		
Matches 363;	Conservative 163;	Mismatches 402;	Indels 390;	Gaps
Qy	872	TATVTVDEANQTSKVAYDVNVDEKTIETLTGNGKKQLQGVKT-----IKLTETSTNGNA	924	
Db	41	talsataeeannntsvtnglnaygdcnfntntnsiadlekhhvqdaykglmlnlekdtt--nk	98	
Qy	925	TTTSTDDHALVKASDIAGNLNLAEETHTTKGTANTALQFTTVYKVVDENDKADDTNNAIT	984	
Db	99	ssflvadntaat-----vgnlrklgwvsskngrtne--ksyqvqkqade-----	140	
Qy	985	VGKDGTSGBKVNTLUKKGKNGLDI--KYDKGTVTFGINTQSGLKAGDSTLTNNGLSIKNT	1043	
Db	141	-----vlfsgsaatvsssskdgkhtitsvktgsfaevktdatg-----	182	
Qy	1044	ASNEQIOVGAGCVKFAWVNGVVVGAGIDGTTTRTDEICFTGTNGSLDKSKPHLSKDGIN	1103	
Db	183	-----gvnadrgkvkaeden--gadvd-----	202	
Qy	1104	AGGKKITNIQSGETAKNSHDVATGGK IYDLKTELENKTSSTAKTAQNSLHFEFSVADEQGN	1163	
Db	203	---kkvatv--dvakaındaatfvkvestdddiengagknettdqal-----	246	
Qy	1164	NFTVSNPSYSDRSKTSDDVITTFAGENGITTKVNGK-----VVRVGIDQTKGLTTPKLTV	1217	
Db	247	-----kagdtitlkagknlkakldqngksvtfalakdlvgtsaakvsdklsi	292	

non-typeable Haemophilus influenzae; antinflammatory; auditory;
antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
diagnosis; immunogenic; antigen.

Haemophilus influenzae.

WO200055191-A2.

21-SEP-2000.

16-MAR-2000; 2000WO-CA00289.

16-MAR-1999; 99US-0268347.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-618897/59.

N-PSDB; AAA92493.

Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Haemophilus influenzae infection

Claim 1; Fig 18; 275pp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia) protein from the non-typeable Haemophilus influenzae (NTHi) strain 33. Hia genes and proteins have antinflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher amount of recovery than a full-length protein.

Sequence 1002 AA;

Query Match 9.2%; Score 952; DB 21; Length 1002;
Best Local Similarity 28.5%; Pred No. 2.2e-34;
Matches 324; Conservative 138; Mismatches 319; Indels 356; Gaps 41;

QY 1084 TGTNGSLDKSKPHLSKDGINAGGKKITNTQSGEIAKNS-----HDVATCGKIYDLKTE 1136
Db 55 tgn-----slnvygknnsfn-----annsladlnkgnsv-----ydglln 93
QY 1137 LENKISSTAKTAQNSLHESVADQ-----GN-----NFTVSNPYSSYDTS---KTSVDVIF 1185
Db 94 Inekgtksk-----flvadettatvgnlrklgwwvstknstkeesngvqkadevlf 145
QY 1186 AGENG---ITTKVNKGVVRVIGIDOTKGLTTPKLTVGNN-----NGKGIV----- 1227
Db 146 egkgdvvtvtsksengkhtvtfalndlnvknatvsklslgangkvvkdtstdanglkfak 205
QY 1228 ---NSQNGQNTITGLSNPLANVTNDKSGSVRTTEGGNIIRDEKTRAASIVDVLISAGNLQ 1284
Db 206 qgtngqngvnhlgiastldldprvggkthltke---isdternraasvqdvlnagwnlr 262
QY 1285 -----GNGEAVDFSVYDVTNFEANGNTTTAKVTDYDTSKTSKVVDVNVNDDTTIEVKDK 1338
Db 263 gaktiggtcvndvstvydtvefasganansvttddnkktt-----vrvdvtgplpvq-- 315
QY 1339 KLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSLGDIQTAKGASQANNSA 1398

KW non-typeable Haemophilus influenzae; antinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX Haemophilus influenzae.
OS WO200055191-A2.
XX 21-SEP-2000.
XX 16-MAR-2000; 2000WO-CA00289.
PF 16-MAR-1999; 99US-0268347.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
PI WPI; 2000-618897/59.
XX N-PSDB; AAA92493.
DR Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for
XX use as antigens and vaccines and for treating Haemophilus influenzae
XX infection
XX Claim 1; Fig 18; 275pp; English.
PS The present sequence represents a Haemophilus influenzae adhesin (Hia)
XX protein from the non-typeable Haemophilus influenzae (NTHi) strain 33.
CC Hia genes and proteins have antinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX Sequence 1002 AA;
SQ Query Match 9.2%; Score 952; DB 21; Length 1002;
Best Local Similarity 28.5%; Pred No. 2.2e-34;
Matches 324; Conservative 138; Mismatches 319; Indels 356; Gaps 41;

QY 1325 DYNVDDTTIEVKDKKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGD 1384
Db 302 -rvdvvgipvq----- 312
QY 1385 IQTAKGASQANNSAGYVDADGNKVIYDSTDNKKYQAKNDGTIVDKTEYAKDKL----- 1437
Db 313 -----yvtdgetv--kvgneyyeakqdsadmkkvengklaktkvkl 355
QY 1438 -----VAQQTDPDTLAQNVKSVINKEQVNDANKKGINEDNAFV-----KG 1480
Db 356 vsangtnpvkisinvadgt---entdavsfkqlkalqdkvltlsasnayangsgdadgkg 412
QY 1481 LE-----KAASDNKT-----KNAAVTVGD----- 1499
Db 413 iqltlnlgnlfnkfstgdgellnkaendvtfpkpgksgvgdgdgkatidgaktgttgive 472
QY 1500 -----LNAVQTPPTTFAGDTGTTAKKLGELTTLTKGGQTDNKLTDNNIGVVAGTGDFT 1552
Db 473 aselvdslnklgkvwgtgtdgtgvtgdthdtl-vksgdvltlkagdnlkvkqegtn-ft 530
QY 1553 VKLAKDLTNE-NSV-----NAGGTKIDKIGISFVDANGQA-----KANTPVLISANCL 1598
Db 531 yalldeltdvksvefkdtagangastktkdglittpangagaagantantistvktkdg 590
QY 1599 DLGGKVISNVGKTKTDANVQQLNEVRNLGLGNDNADGNQVNIADIKDPNSGSSN 1658
Db 591 sagnkavknvsglikkfgdanfopltssadnlctkydynaykgtlnldekskgtqptvad 650
QY 1659 RTVIKAGTVLGGKGNNDTEKLTATGGVGVGDKDGNGANGDSLVN-WYTKQDKSGSKALLAT 1717
Db 651 nt-----aatv-----gdrlgigwv-isadktkgeinke 678
QY 1718 YNAAGQTNVYTNPAEALDRINEQGIREFPHVNDGNOEPVQGRNGIDSSASGK-----HSV 1773
Db 679 yna-----qvrnanevfkfsg-ngl--nvsgktldngtr 709
QY 1774 AIGFOAKADGEAAVAIGRQTQAGNQSIAGDQNAQATGDSIAIGTGNVAGKUSIGAIGDP' 1833
Db 710 eitfelakdena-----lafgsgskalrdntvaigtgnvnaeksgafgdp 755
QY 1834 STV---KANDSVSGVNNQOFTDATQDVFVGVCNNI-----TVTESNSV----- 1873
Db 756 nyiedkaggsyafgndnri---tskntfivgnsnkrangnvltekeevvkgdgaktk 812
QY 1874 -----ALG---SNSAISAGTHAGTQAK-----KSDGTAGTTTACATCTVKGFAQTAVGA 1921
Db 813 vtvpgalgetvsnvylgnastatkdkgnlksgdtagnnttttagatgtvngfagatahga 872
QY 1922 VSVGASGAEERRIONVAAGEVSATSDAVNGSGLYKATQSIANATNELDHRITHONENKANA 1981
Db 873 vsvgasgeerriqnvaageisatstdaingsqlyavakgvcn---lagqvknvkgkrada 928
QY 1982 GTSSAMMASMPQAVTPGRSMVYTGCIATHGOGAVAGVLSKLSNDNGQWVVFKINGSADTQG 2041
Db 929 gtasalaasqlpqasmpgkmsvsiagssyqgsglaigvrsldngkviirlsgttnsqg 988
QY 2042 HVGAAGVAGGFHF 2053
Db 989 ktgvaagvgvqw 1000

RESULT 12
AAB23854
ID AAB23854 standard; Protein; 1002 AA.
XX
AC AAB23854;
XX
XX 17-JAN-2001 (first entry)
DT
DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain 33.
XX
XX Hia; adhesin; Haemophilus influenzae adhesin; NTHi; Infection; vaccine;

Db 316 ----- 315
Qy 1399 GYVDADGNKVIYDSTDNKYYQAKNDGTVDKTEKAVAKDLV---AQATPDGTLQAMNVKYS 1455
Db 316 -yvtedskttvv---kvneyyeakqgdmdkkgvengklaktkvklvsangt-npvklsl 371
Qy 1456 VINKEQVND-----NKKOGINEDNAFVKGLEKAAASDNKTKNAAVTVG-DLNVA 1504
Db 372 vadgtdtdavsfkqkalkqdkvtlsasnyang---gsdadgkatqtlgndlnfkf 427
Qy 1505 QTP-----LTFAGDTGTAKKLGET-----LTIKGGQDTNKLTD-----NNIG 1543
Db 428 kstdsellnikaaadtvtftpkksqvqvgdgdgkatidgakttgilveaselsvlnklg 487
Qy 1544 VVAG-----TDG-----FTVKLAKDLTNLSV- 1565
Db 488 wkvgvgdgtgatdgttdtlvksqdkvtlkgdnlkvkqegtnftvylrldeltgkvsve 547
Qy 1566 -----NAGGTKIDEGISFV---DANGOA---KANTPVLISANGLDLGGKVISNVGKGT 1612
Db 548 fkdtdengangastkitdglittpandangaaatdadkikvasdgisagnkavknvsgl 607
Qy 1613 KDTDAANVQQLNEVRNLGLGNDAGNQVNIADIKDPNSGSSSNRTVIRKAGTVLGGK 1672
Db 608 kkfgdanfnltssadnltkydnaykgtlnldkskqkqtpvtadnt---aatv----- 659
Qy 1673 NNDTEKLTATGQVGVQVDKGNANGDLSNV-WV---KTQDKGSKKALLATYNAAGQTNVYTN 1729
Db 660 -----gdrlrglvisadkktgeskeysaqvrnanevkfksg 696
Qy 1730 NPAEALDRINEOGRIFHHVNDGNEPVGGRNGIDSSASGKHSVAIGFOAKDGEAAVAI 1789
Db 697 nginvsgktldngtr-----eitfelakdena----- 723
Qy 1790 GROTQAGNQSIAGIDNAQATGDSIAIGTNVWVAGKHSAGIDPSTV---KADNSYSVGN 1847
Db 724 -----lafsgskalrdntvaigtgnvnaeksgafgdpnyiedkaggsyafond 773
Qy 1848 NOFTDATQDVFVGNN-----ITV-----TESNSVALG 1876
Db 774 nrit-skntfvlngvnakykangdvdttvtvkdkdgtttvtpkalgatvensvylg 832
Qy 1877 SNSAISAGTHAGTQAKKSDGTAGTTTATGATGTVKGFAGQATVAGVAGSVGASGASERRIONV 1936
Db 833 nk-----statkdkgnksdgcagntttagttgtvngfagatahdavsvgasgeerriqn 889
Qy 1937 AAGEVSATSDAVNGSOLYKATQSIANATNELDRIHQENKANAGISSAMAMASMPQAY 1996
Db 890 aageisatstdaingsqlayavakvtn---laggvkvgradagtasalaasqlpqas 945
Qy 1997 IPGRSMVTGGIATHNGQGAVALGSLSDNGQWVFKINGSADTQGHVGAAGVAGPHF 2053
Db 946 msgksmvsiaqsgyqggsglaigvrsisdngkviirlsgttnsqgktgvaagvgvqw 1002

RESULT 13

AAW56322 standard; Protein; 2039 AA.

AC AAW56322;
AC AAW56322;
DT 19-AUG-1998 (first entry)
DE Haemophilus paragallinarum antigenic protein #2.
KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW vaccine; chicken infectious coryza; CIC; fowl.
XX Haemophilus paragallinarum.

OS key Location/Qualifiers
FH Peptide 1..70
FT /label= signal

FT Protein 71..2039
XX /note= "antigenic protein"
PN WO9812331-A1.
XX 26-MAR-1998.
XX 12-SEP-1997; 97MO-JP03222.
XX 19-SEP-1996; 96JP-0271408.
XX (KAGA) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
XX (KAGA) CHEMO-SERO-THERAPEUTIC RES INST.
XX Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX WPI; 1998-230318/20.
XX N-PSDB; NAAV22837.
XX Antigenic polypeptide from Haemophilus paragallinarum induces HI
XX antibody production - and is useful for diagnosis of and preparation
XX of vaccines for chicken infectious coryza
XX Claim 5; Page 71-87; 108pp; Japanese.
XX The present sequence represents an antigenic protein derived from
XX Haemophilus paragallinarum strain C-53-47. The antigenic protein
XX stimulates the production of HI antibodies in fowl. The protein
XX and DNA coding for it can be used in the preparation of vaccines
XX for the prevention of chicken infectious coryza (CIC). The protein
XX and its antibodies can be used in the diagnosis and treatment of CIC.
XX Sequence 2039 AA;

Query Match 7.8%; Score 809; DB 19; Length 2039;
Best Local Similarity 20.3%; Pred. No. 1e-27;
Matches 501; Conservative 311; Mismatches 810; Indels 848; Gaps 101;

Qy 1 MNHIYKVFNFKATGTFMAVAYAKSHSTGGSCATGQVSVRTLSFARIAALAVLV---I 57
Db 1 mnkvfkkyvkvqmivvselannkktasknhttaffql-ftktylallinal 59
Qy 58 GATL-----NGSAYAQQITTKIEIGQTNKIN-----NTLKGDMAL----- 91
Db 60 gtslfpqlanakfllevynsvklqhvnsqvpdsdv- lnpssgenvgmnsdgvagtrga 118
Qy 92 -----ATGEASTAFGSLSKAQSQAIAIGS----- 116
Db 119 vnnysatgsiaigqgakndnwastrsiaigqaknesiasdsvalsnalnrfkksivigl 178
Qy 117 --VKPDPNNGSNGN-----VGSNAK--GNE-----SIAIGDVLAEGDASTA 154
Db 179 tytqlprapesrgsvvigenaksgnqsvslgnawsktnsisagagtfaegkstia 238
Qy 155 IGSDDL-----LPKN--LDLKNFEH--KLINGHEILK----- 184
Db 239 igtdkiltgntnyndklpapswdgtrgtkaptnsiwdifselymgkktngtdydkndrdpn 298
Qy 185 ----IQTSTDGKIY-----RRTRAQGHASTAVGAMSVAQGHFSN 220
Db 299 kpeafytsdfksryvnnpstptyaaklgalaalgsrtlaagemstavgslafaladkt 358
Qy 221 AFG--TYATAEAAYSLAVGLAAQATKQSSIAVGSNAKAN-----AFAATAIGGNTVVN 271
Db 359 amglrfsfvakdvagtaigeesrtfakdsvalgnkteasnaagsmaygykakavagagaia- 417
Qy 272 LGRGVALG-----FGSOILDRDNNTDASAVVPLGKTLADQYKATRGQSDTDIF----- 319
Db 418 igaevaagaefdsdsgagnlllnr-----gayatlks--akdsddikagdainvttqffd 469
Qy 320 -----SIGNSNNN-----NSSIRRKIIINVGAS--- 342

Db 470 nmltqshlytyentyltsagdiakktlaavdgdknaiaignktafskansvalgsval 529
 Qy 343 -RDTDAVNAQLKLYEEL-ANRKITFKDGDNNNSVERG-----LGNLTITIK 388
 Db 530 asagnafalgysilvaplantivlgvygagysfvggswstlsartvlgysasis 589
 Qy 389 GDAQTNALTEANIGVYTDGNGLKVKLAKELTGLTSVSATNKITVNTNNNAELSGGLT 448
 Db 590 sdsdlsiamgvnafi--gng-----snssialg--- 615
 Qy 449 FSPITGKTKDVTYISIDGLKFTNDNSNIATKGTTRITKKKIGFAGTNDGV----- 498
 Db 616 ----tgsiaknakspdsialgkds-----ridakldngvlytpqvydet 657
 Qy 499 -----DESKPYLDNBLKLVG-NSTLNSGLTVNNTTKOIQVGANGIKFATVANNVA 550
 Db 658 trafrtdenkdm-rqamalgfnakvsgkgkme--tginsmaigar--sqatlgntsa 712
 Qy 551 NTSATVGTARITEKIGFAGTNDGV-----EQAPYLDKERL-----KVGREVEITDS 598
 Db 713 -----lgvnaktdytweqleapdwskgaistskigvisv--- 750
 Qy 599 GINAGNKITGLTNGIANTDAVTKOLK----- 626
 Db 751 gskgserrlvnvasgsldtdavnavaqiktieerfgseidllqngggvqylsvektninge 810
 Qy 627 -----DAKPTLNAG--DGISIN-----SNNGD 646
 Db 811 agrvasqirgeskryvkvikltqilyidarkklngefdqtsidkiskavqealeaaysge 870
 Qy 647 LVDSSGNITTFYINLVKTKLNSNGTSGNKFESVNAHDNNS-----LVTAKD 695
 Db 871 lkttaselnrvamqlnaett-vnfdgfkngyktqienatnadseknvgglsppqvialka 929
 Qy 696 LADYLNKVNETADSLPBFKQVQNDNSN-----NAITVGDGTNGKTFNTLKLGENCVNI 750
 Db 930 nnnylndgagqgdsiafagqaktseannaglagkqalaifqans-----saenaisi 981
 Qy 751 TTN-----RATGTVTFIDGOSNGLT-----PKLTVGSDTNGNRLVIEOVPSADGNS-TPKNI 801
 Db 982 gtnsdtsmtgvalg-----kgatvtagkpsiaigqdstvansaisrtssvmlngltfnn 1037
 Qy 802 IKGLSPTLPSIA-SPSGR-----NIALGN-----TIEEKDKNAAISIDVLN-AGFNLKN 850
 Db 1038 fagspetlgvlsigtagkerkivnvaagdisqtsteaingsqlyatnfmnlkagsvktn 1097
 Qy 851 GKDKQFVSTYDVFIDGNATATVTVYDEANQTSKAVDVNVDEKTIETLTDGNCCKQLGV 910
 Db 1098 fggnanlatdgtitfnigtgqgdtlndain-----nvltklslateeeevsvg 1148
 Qy 911 KTIKLTETSTNGNATTFSTDDHALVKASDIAGNLNTLAEETHTTKGTANTALQTFVKK 970
 Db 1149 eav---ydalkgaktvsaeankgiglvdvvkans---pitvepstdnkkkftvyl 1202
 Qy 971 VDENDKADDTNATVQKDG-----TSKGVTYTLKLGKGLDKTKDGVTFGINTQSGL 1025
 Db 1203 mkd---legvnsitfdksqgdlnqvtgrmsagltfkkq---dttngsttf---aedgl 1253
 Qy 1026 KAGDSTT-----LNNGLSIKNTASNEQIQVGADGVKFAMVNGVVGAGIDGTTTRI 1076
 Db 1254 ti-dstlnsaqalnkvsvrdsfsvkn-----gsdesklaskls-igaenaehvev 1302
 Qy 1077 TRDEIGFTGNGSLDKSKPHLSKDGINAGK-----KITNTOSGEIAKNSHDVTGKI 1130
 Db 1303 tksgiaiakadnts-dkssitlagdaitlagnatgtaitkigtvadgnitvnskdavnngql 1361
 Qy 1131 YDL-----KTELENISSTAKTAQNSLHEFSVADQGNFTVSNPYSYDTS 1177
 Db 1362 rtlilgvsagkiggttekttiiseaisdvkqaltatdaykadmknk-----lv 1409
 Qy 1178 KTSOVIYTFAGNGITTKV-NKGVVRVCI-DOTKGLTT-----PKLTVGN 1220
 Db 1410 kltgdglntstnidasvedngvvkftikldltgltklatieslnasqnliaagctvtvgge 1469

Qy 1221 NGKGIIVNSQNGONTITGLSNTLANVTNDKGSVTRTTEQGNIIKDEKTRAAISIVDVLSSG 1280
 Db 1470 t-egivltksgsgndrtlsisgagnaatdgikvsgvkagt--adtdavnkgl-dklfka 1525
 Qy 1281 FNLQNGEAVDFVSTYDVTNPFANGNTTTAKVTVDDTSKTSKVYDVNVDDTTIEYK---D 1337
 Db 1526 ln-----dalgttdlavtknpqts---ifnplngtapttkfkdavdklktavntgvg 1574
 Qy 1338 KKLGVKTTTITSTGCTGANKFALPNOATGDALVKASDIAHLNLTLSGDIQTFAGASOANNS 1397
 Db 1575 skvgillatgldgldagnk--isnvadgd-----ispsgsvvtlgr----- 1613
 Qy 1398 AGYVDADGNKVIYDSTDNKYQAKNDGTVDKTEVAKDKLVAQAQTPDGTGLAOMNVKSVI 1457
 Db 1614 -----qlyalmqkgirvygdevsptktgttapt----- 1642
 Qy 1458 NKEQVNDANKKQGINEDNAPVKLEAKASDNKTKNAAVTVGDLNAVAQTPLTFAGDTGTT 1517
 Db 1643 -----sstqggattantagvavapagnvatgdiapt 1672
 Qy 1518 AKKLGETLTIKGGQTDTNKLTDNNIGVVAGTDPGFTVKLAKDLTNLNSVNAGGPKIDEKI 1577
 Db 1673 qpalpemt-----alvgdhlavplpg---slkghdnhvkttisag-----ndvgi 1716
 Qy 1578 SF-----VDANQAKANTPVLNGLDLGGKVISNVGKGTOTDAANVQQLNEV 1626
 Db 1717 siqpnisiennlivigsnkpekaklaaqegnal-----vitn-----kddgnaamv--fnne 1765
 Qy 1627 RNLLGLGNDNADGNQVNIADIKKDPNS---GSSNRRTVIKAGTVLGGKGNNDTEKLATGG 1683
 Db 1766 kmnlvl-----sdkkakpravldgngqaltiv-----gnddsq--vlls 1802
 Qy 1684 VQGVVDKOGNANGDLSNVWVKTKDCKSKALLATVNAAGOTNYVTNNPAEALDRINEOGI 1743
 Db 1803 skkgkldign---disrlsvtttert-----nadgqlekvetfsatmdd-----gl 1844
 Qy 1744 RFPHVNDGQEPVQGRNGIDSSASGKHSVAIGFOAKKADGEAAVAIGRQTOAGNOSIAIG 1803
 Db 1845 kf-----kadgdkvi-----nkkln 1859
 Qy 1804 DNAQATGDGSIAGTGNVAGKHSAGIDPSTVKADNSYSVGNNOFTDATOTDVFVGVN 1863
 Db 1860 etvelvgdenvt-----tsitddnkvkslnkkla-----idevklpn 1897
 Qy 1864 NIIVTESNVAIGNSAISAAGTHAGTQAKKSDCTAGTTTTAGATGTVKGFACQTAGVAVS 1923
 Db 1898 tdpdaqgdsivlnnggihagnkvilgvtkasd----- 1929
 Qy 1924 VGASGAERRIQNVAAAGEVSATSTDAVNGSQLYKATOSIANATNELDHRHONENKANAGI 1983
 Db 1930 -----dptsavnrqqlntvldnqnnfnqvnqrldltresragl 1969
 Qy 1984 SSAMAMASMPQAVIPGRSMVTGGIATHNGGAVAGVGLSKLSDNGQWVFKINGSADTQGHV 2043
 Db 1970 agamataslgnvalpgkttisvtatfkgenavaimrslsdngkvglrlsgmstngdk 2029
 Qy 2044 GAAVAGAGFHF 2053
 Db 2030 gaamsvgftf 2039

RESULT 14
 'AAW56319
 ID AAW56319 standard; Protein; 2042 AA.

XX AAW56319;
 AC AAW56319;
 XX 19-AUG-1998 (first entry)
 DE Haemophilus paragallinarum antigenic protein #1.
 XX Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
 KW

1103 NAGGKITTNIQSSEIAKNSHDAVTGKIIYDLKTELENKISSTAKT-AQNSLHFSVADEQ 1161
1162 GNNFTVSNPSSYDTSK---TSDVITFAGENGITTKVNGVVRVIGDQOTGLITPKLTV 1217
1422 --klttskvesytnkfanfstdiillsdgrsnattandvgkrlsdgftiksesntfl 1479
1218 GNN--NGK---GVINSQNG---QNTITGLSNTLANVNDKGSVRTTEOGNIKDEDKT 1268
1480 gskayngsdslgvmvdydngvfxlismtalttslan-----tfakidasnltdsne 1533
1269 RAASIVDVLGAGENLQNGEAVDFSVSYDTVNFANGNTTTAKVTVYDTSKTSKVYDVNV 1328
1534 kwtalnvs-----ktevdaelqkskvtl---t 1559
1329 DDTTIEVKDKKLGVKTTTLSTGTGANKFALSNOATGDALVKASDIVAHLNLSGDIOFA 1388
1560 pdsqlifatkgag---sgnnagidagnkkslnvadg-----isptsdgvvtg 1604
1389 KGASOANNSAGYVDADGNKVYIDSTDNKYYQAKNDGTDVTKTEVAKDKLVAQAQT---PD 1445
1605 r-----qlvalmqkgirvvygdevspktktqtaptanpt 1638
1446 GTLAQMNVKSVINKEQVNDANKOGINEDNAFVKGLEKAASDNKTNAAVTVGDLNAVAQ 1505
1639 at-----taptasstggwattantagvav 1663
1506 TPLTFAGDTGTAKKLGETLTKIGQDQDTNKLTDNNIGVAVGDTGTVKLAKD---LTNL 1562
1664 agnavtdgiaptcptlpemt-----alvdhlaavlpgg---sikihgdhvktti 1711
1563 NSVNAVGGTKIDEGISF-----VDANGAKANTPVLISANGLDLGGKVISNVGKTKDQDA 1617
1712 sadnqvgsil-qpnisiennlvisndpekaklaaqegnal-----vitn-----kddgna 1761
1618 ANVOQLNEVRNLLGLGNDNADGNQVNIADIKKDPNSGSSNRVTIKAGTVLGK----- 1671
1762 amv--fnneknmvlisdkea-----kprvllidgngaltl 1794
1672 -GNNDTEKLATGGVQGVQDKDNGDLSNVWVKTKDGSKKALLATVNAAGQTNVYTN 1730
1795 vgnddsg--vtlsskkgkldgn---dlrlsvtttert-----nadgglekvets 1839
1731 PARAIIDRINEQTRFRHVNNDQEPVQVQGRNGIDSSASKHSVAIGFOAKADGEAANAIG 1790
1840 fatumd-----klf-----kadgkvi--- 1857
1791 ROTOAGNQSIAGDQAATGDSIAITGTGNVAGKHSAGIDPSIVKADNSVSYGNNQOF 1850
1858 -----nkklnetvelvgdenvt-----tsitdnenkvkslnkki 1891
1851 TDATQDVFVGNNTVTSNSVALGNSAISAGTHAGTQAKKSDGTAGTTTAGATGTV 1910
1892 a-----idevktpdpdaqkdsivimngghagnkvilgvkasd----- 1932
1911 KGFAGQTAAGVAVSVGASGAERRIQNVAAGEVSATSDAVNGSOLYKATOSIANATNLDH 1970
1933 -----dptsavnrnglntvldnvgnnfnqvnq 1959
1971 RIHONENKANAGISSAMAMSPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSNDGQWV 2030
1960 rigdltrerasragiagamaslqvalpgkttisvgtatikgenavaimsrslsdngkv 2019
2031 FKINGSADTQGHVGAAGVAGPHF 2053
2020 irlsgmstsnkgkaeamsvgfsf 2042

RESULT 15
AAB23858
ID AAB23858 standard; Protein; 1094 AA.
XX

AAB23858;
17-JAN-2001 (first entry)
Haemophilus influenzae adhesin (Hia) protein from NTHI strain K9.
Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
non-typeable Haemophilus influenzae; antiinflammatory; auditory;
antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
diagnosis; immunogenic; antigen.
Haemophilus influenzae.
WO200055191-A2.
21-SEP-2000.
16-MAR-2000; 2000WO-CA00289.
16-MAR-1999; 99US-0268347.
(CONN-) CONNAUGHT LAB LTD.
Loosmore SM, Yang Y, Klein MH;
WPI; 2000-618897/59.
N-PSDB; AAA92497.
Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
use as antigens and vaccines and for treating Hemophilus influenzae
infection
Claim 1; Fig 22; 275pp; English.
The present sequence represents a Haemophilus influenzae adhesin (Hia)
protein from the non-typeable Haemophilus influenzae (NTHi) strain K9.
Hia genes and proteins have antiinflammatory, auditory and antibacterial
activities, and can be used in the production of a vaccine. An
immunogenic composition comprising an Hia gene, a polypeptide encoded
by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
protection against disease caused by Haemophilus strains in a
susceptible host, preferably a human. An Hia protein is useful as an
antigen, in immunogenic preparations including vaccines, as a carrier
for other immunogens, and in the generation of diagnostic reagents. Hia
is useful for treating diseases caused by the infection of Haemophilus
influenzae such as meningitis, epiglottitis, septicaemia and otitis
media. Recombinant production of Hia favours high recovery of the
protein compared to the low recovery of native protein from Haemophilus
influenzae species. A truncated protein has a significantly higher
amount of recovery than a full-length protein.
Sequence 1094 AA;
Query Match 6.9%; Score 710; DB 21; Length 1094;
Best Local Similarity 25.1%; Pred. No. 1.1e-23;
Matches 315; Conservative 162; Mismatches 438; Indels 340; Gaps 56;
QY 953 HTTKGTANTA---LQTFVKKVDENDKADDTNAITVGKDTGKVTLLKLGKNGL-DIK 1008
Db 26 htkrasatvatavlatqlsataeanssasvtsrlv-----ygdtnktkfaannsladin 80
QY 1009 TDKGDTVFGINTOSGLKAGDSITLNNNGLSIKNTASNEIQVGDGKVFAMVYNNVYGA 1068
Db 81 kqndg-----vhdgl-----lnhenganckllvddntaalvgdlrklgwvstknkg 128
QY 1069 GDGTTRITR-DEIGFTGTNGSLDKSPHLSKDGINAGGKITTNIQSGETAKNSHDAVTC 1127
Db 129 eneksqqvkqadevlfksgk-----gvqvtstscengkhait- 165
QY 1128 GKIYDLKTELENKISSTAKTQAQNSLHFSVADEQNNFTVSNPSSYDTSKTSKVITFAG 1187
Db 166 ---falakaldmr-----tatvsdtlttgg 187

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:37:49 ; Search time 44.72 Seconds
(without alignments)
945.259 Million cell updates/sec

Title: US-09-361-619-9
Perfect score: 10356
Sequence: 1 MHHYKVFINKATGTFMAVA.....NGSADTQGHVGAAGAGFHF 2053

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7357	71.0	2123	4	US-08-968-685A-10
2	1206	11.6	2353	4	US-09-377-155-33
3	1206	11.6	2353	4	US-08-913-942-4
4	890	8.6	1912	1	US-08-409-995-4
5	890	8.6	1912	3	US-08-685-467-4
6	690	6.7	1098	1	US-08-409-995-2
7	690	6.7	1098	3	US-08-685-467-2
8	690	6.7	1098	4	US-09-377-155-32
9	690	6.7	1098	4	US-08-913-942-2
10	583.5	5.6	679	4	US-08-913-942-15
11	546.5	5.3	1600	2	US-08-617-697-10
12	515	5.0	1599	2	US-08-617-697-9
13	512	4.9	1529	2	US-08-728-470-10
14	512	4.9	1529	4	US-08-719-641-10
15	469	4.5	1477	1	US-08-038-682-4
16	469	4.5	1477	1	US-08-302-832-4
17	469	4.5	1477	2	US-08-530-198-4
18	469	4.5	1477	2	US-08-469-880-4
19	469	4.5	1477	2	US-08-728-470-4
20	469	4.5	1477	2	US-08-617-697-4
21	469	4.5	1477	4	US-08-719-641-4
22	459	4.4	1338	2	US-08-728-470-9
23	459	4.4	1338	4	US-08-719-641-9
24	456	4.4	1536	1	US-08-038-682-2
25	456	4.4	1536	1	US-08-302-832-2
26	456	4.4	1536	2	US-08-530-198-2
27	456	4.4	1536	2	US-08-469-880-2

28	456	4.4	1536	2	US-08-728-470-2	Sequence 2, Appli
29	456	4.4	1536	2	US-08-617-697-2	Sequence 2, Appli
30	456	4.4	1536	4	US-08-719-641-2	Sequence 2, Appli
31	440	4.2	1612	1	US-08-169-927-2	Sequence 2, Appli
32	393	3.8	599	4	US-09-377-155-15	Sequence 15, Appli
33	391	3.8	592	4	US-09-377-155-17	Sequence 17, Appli
34	388	3.7	594	4	US-09-377-155-9	Sequence 9, Appli
35	387	3.7	598	4	US-09-377-155-5	Sequence 5, Appli
36	385.5	3.7	592	4	US-09-377-155-2	Sequence 2, Appli
37	385	3.7	598	4	US-09-377-155-13	Sequence 13, Appli
38	384.5	3.7	658	1	US-08-409-995-5	Sequence 5, Appli
39	384.5	3.7	658	3	US-08-685-467-5	Sequence 5, Appli
40	384.5	3.7	658	4	US-08-913-942-5	Sequence 5, Appli
41	382.5	3.7	589	4	US-09-377-155-19	Sequence 19, Appli
42	381.5	3.7	591	4	US-09-377-155-11	Sequence 11, Appli
43	381.5	3.7	591	4	US-09-377-155-21	Sequence 21, Appli
44	377	3.6	594	4	US-09-377-155-7	Sequence 7, Appli
45	353	3.4	607	1	US-08-409-995-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968 685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 71.0%; Score 7357; DB 4; Length 2123;
Best Local Similarity 70.6%; Pred. No. 0;
Matches 1571; Conservative 120; Mismatches 261; Indels 272; Gaps 34;

QY 1 MHHYKVFINKATGTFMAVAEYAKSHSTGGSCATQGVSRVTLSPARTAAALVIGAT 60
|||||

Thu Sep 13 14:19:21 2001

Db 1 MNHUYKVENKATGTFMAVEYAKSHSTGGSCATQGVSGVRLTSFARIAALAVLVIGAT 60
 QY 61 LNSGSAVOITTKIEGTQNTKNNTLKGDAATGEASTAFGLSKAOGSOAIAIGSVKPD 120
 Db 61 LNSGSAVOITTKIEGTQNTKNNTLKGDAATGEASTAFGLSKAOGSOAIAIGSVKPD 111
 QY 121 PNGSGNNGVSHAKGNESIAIGDVLAEADGASIAIGSDDLVLPK-NLDLKNFHKHLHGH 179
 Db 112 HNSNNNANIGAKAGNESIAIGDVLAEADGASIAIGSDDLVLPK-NLDLKNFHKHLHGH 171
 QY 180 EILKIKIOTSTDGKI-KYRTRAGHASTAVGAMSTAQGHFNAFGTYATAEAYSLAVGL 238
 Db 172 KALNDIYQLADNLQYRTHRAOCHASTAVGAMSTAKGHFNAFGTYATAEAYSLAVGL 231
 QY 239 AAOATKOSSTAVGASNAKANAFAATAGGTVNVLGRVALGFGSQTLDRDNDNDAS--A 295
 Db 232 TATAKAASSIYVAGSNAOIAIGFATVAGGTOVNLNRGIALGFGSQVQLQKNDVNAANVRA 291
 QY 296 YVP-LGKTLADQYKATQCGDSTDFISGNSNNNSIRKLIINVAGAGSRDIDAVNVAOLK 354
 Db 292 YAPDNDQPIDNRYKATEFKNGATDVFISGNSGND-SIRKLIINVAGAGSRDIDAVNVAOLK 350
 QY 355 LVEELANRKTFFKGGDNNNSVERGLNTLTIKGDAQTNALTEANIGVVYDGLKVLK 414
 Db 351 EAVRLANROITFKGDDSN--NRVEKGLKLTITVGGNAQTSALTDHNLGVVQNGDGLKVQL 408
 QY 415 AKELTGLTSVATNKTIVTSNTNNNAELQSGGLTFSPITCTKTKVYSIDGLKFTNDSN 474
 Db 409 AETLSLKWVTEN-----LT-----AN 426
 QY 475 SIATGTRITRKKKIGPACTNDGVDESKPYLD----- 506
 Db 427 EKVTVGKRLTLDKIGFTNDMNGIDESKPYLDKDTGTHAGGQKTKLTAGVWDDAATYG 486
 QY 507 ----- 506
 Db 487 QLKVNQTAESALQTFVKKVKNNDANDANSKIITVGNKNNKPDGTQVNTLKLGENGVGV 546
 QY 507 -----NEKLKVGNSLNSGLTVNNNTGNKQIOVGANGIKFATVANNVANT 552
 Db 547 TTETNGTVFGLNONGNLTGVNSTLNDGLSVKNTNSKQIOVGADGITTIDISNKP- 605
 QY 553 SATVGTARITEKIGFAGTNDGVDEQAPYLDKERLKVGRVEITVDSGINAGNHKLTGL-- 610
 Db 606 AGENTRITRDGIGFAGTNDGSLDANKPRL-----TPTGINAGGKELTNVQS 652
 QY 611 -----TNG-----TANTD-----AVTIKQLAKDAK--PTLNAKD--GISNSNGDLV 648
 Db 653 AINPATNGGQLDPMNRLSTANTKSGSAATIKDLNLNLSQVPLTFAGDTGPNVTKLGEI- 711
 QY 649 DSSGNITPTPTYNISVTKTLNLSNGTSNNKFSVSNADHNSLV-TAKDLADYLNKVN- 705
 Db 712 -----LKVGGKTADDLTKNIGVADSTNLSVLAKTLSL-LDAVNTKT 758
 QY 706 -----TADSLPSPKQVQNGSNNAITVGDNTGKT--FNTLKLKNGVGNITNRA- 755
 Db 759 LTASDKVTVDGSGNNAKTALONG-----LTFKONTGATPATNSKTIYGVGDKLFTDNGI 813
 QY 756 --TGTV-----TFGIDQSG--LTPKLTVGS--DTNGNRLVIEQVPSADGNS 797
 Db 814 ALDGTYYITKDKVGFAGKQDSGLDKSRPYLDKDLKVGVEITNGI-----NA 861
 QY 798 TKNIKGLSPLTSPASPSGRNIALGNTIEEKSNASADDDVLNAGFNLKNNGDKDFV 857
 Db 862 GKATITGLSNTLTATNATTGHVTLQGLVDSDTKTRAASIGDVLNAGFNLKNNGDKDFV 921
 QY 858 STYDTPDFIDGNATATVYDEANQTSKVAIDVNVDEKTIETLC-DNGKQLQGVKTKLT 916
 Db 922 STYDTPDFIDGNATATVYDEANQTSKVAIDVNVDEKTIETLC-DNGKQLQGVKTKLT 916
 QY 917 ETSTNGN-ATTFSTD--DDHALVKAASDIAGNLALAEIHTTKTANTALQTFVKKVDE 973
 Db 980 KTDAGDKAINFVSNVSGDDKALINAKDITADNLNLTAGEIRNTKGTADTALQTFVKKVKE 1039

QY 974 N-DKADDTNALTIVGKDGTSKVNLTLLKGNGLDIKTDKDGTVTFGINTQSGLKAGDSTT 1032
 Db 1040 NGDDDDADADITIVGKDAKNTQVNTLLKGNGLDIQTKDGTGTVTFGINTQSGLKAGNNTT 1099
 QY 1033 LNNGLSIRKNTASNEOIQVAGDGVFAMVNVNGVAGIDGTTRITRDEIGTGTNGSLDK 1092
 Db 1100 LNNGLSIRKNTASNEOIQVAGDGVFAMVNVNGVAGIDGTTRITRDEIGTGTNGSLDK 1159
 QY 1093 SKPHLSKDGINAGGKITTIOGSEIAKNSHDAVVGKIYDLKTELENKISSSTAKTAQNSL 1152
 Db 1160 SKPHLSKDGINAGGKITTIOGSEIAKNSHDAVVGKIYDLKTELENKISSSTAKTAQNSL 1219
 QY 1153 HEFSVADQEGNNTFVSNPYSSYDTSKTSDVITFAGENGITTKVKNKGVVRVGDIDQTKGLTT 1212
 Db 1220 HEFSVADQEGNNTFVSNPYSSYDTSKTSDVITFAGENGITTKVKNKGVVRVGDIDQTKGLTT 1279
 QY 1213 PKLTVGNNGKGVINSONGONTITGLSNTLANVNDKGSVRTTEQGNIIKDEDKTRAAS 1272
 Db 1280 PKLTVGNNGKGVINSONGONTITGLSNTLANVNDKGSVRTTEQGNIIKDEDKTRAAS 1339
 QY 1273 IVDVLSAGFNLOGNEAEVDFVSTYDVFNFANGNTTAKVYDDTSTKTSKVVDVNVDDTT 1332
 Db 1340 IVDVLSAGFNLOGNEAEVDFVSTYDVFNFANGNTTAKVYDDTSTKTSKVVDVNVDDTT 1399
 QY 1333 IEVKDKKLGWVTTTLTSTGTCANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKGAS 1392
 Db 1400 IEVKDKKLGWVTTTLTSTGTCANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKGAS 1459
 QY 1393 QANNSAGVVDADGNKVIYDSTDNKKYQAKNDGTVDKTEKAKKLVAAQATPDGTTLAQM 1452
 Db 1460 QANNSAGVVDADGNKVIYDSTDNKKYQAKNDGTVDKTEKAKKLVAAQATPDGTTLAQM 1519
 QY 1453 YKSVINKEQVNDANKKQGINEDNAFVKLEKASDNKTNAAVTVGDLNAVAQTPLTFFAG 1512
 Db 1520 YKSVINKEQVNDANKKQGINEDNAFVKLEKASDNKTNAAVTVGDLNAVAQTPLTFFAG 1579
 QY 1513 DTGTTAKKLGELTITKGGQTDNTKLTNNIGVWAGTDTGTVVKLAKDLTNLNSVNAAGTKI 1572
 Db 1580 DTGTTAKKLGELTITKGGQTDNTKLTNNIGVWAGTDTGTVVKLAKDLTNLNSVNAAGTKI 1639
 QY 1573 DEKGSFVDANQOAKANTPVLNGLDGLGKVIKSVNGKGTDTDAANVOQLNEVRNLGL 1632
 Db 1640 DDKGVSVFVSSGOAKANTPVLNGLDGLGKVIKSVNGKGTDTDAANVOQLNEVRNLGL 1699
 QY 1633 ---GNDADGNQVNIADIKKDPNSGSSSNNRTVIKAGTVLGGKGNNDTEKLTGQVGVVD 1689
 Db 1700 GNAGNDADGNQVNIADIKKDPNSGSSSNNRTVIKAGTVLGGKGNNDTEKLTGQVGVVD 1759
 QY 1690 KDGNAKDLSNVWVKTKOGSKKALLATYNAAGOTNYVTNNPAAEIDRINEOGIRFFHVN 1749
 Db 1760 KDGNAKDLSNVWVKTKOGSKKALLATYNAAGOTNYVTNNPAAEIDRINEOGIRFFHVN 1819
 QY 1750 DGNQEPVQVRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDQAQAT 1809
 Db 1820 DGNQEPVQVRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDQAQAT 1879
 QY 1810 GDOSIAIGTGNVAGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDQAQAT 1869
 Db 1880 GDOSIAIGTGNVAGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDQAQAT 1939
 QY 1870 SNSVALCSNSAISAGTHAGTQAKSDGTAGTITTAGATGTVKGFAGTAVGAVSVGASGA 1929
 Db 1940 SNSVALCSNSAISAGTHAGTQAKSDGTAGTITTAGATGTVKGFAGTAVGAVSVGASGA 1999
 QY 1930 ERRIQNVAAGEVSATSDAVNGSOLYKATOSIANATNELDHRHIOHENKANAGISSAMAM 1989
 Db 2000 ERRIQNVAAGEVSATSDAVNGSOLYKATOSIANATNELDHRHIOHENKANAGISSAMAM 2059
 QY 1990 ASMPQAYIPGRSMVTGGTATHNGOGAVAGLSKLSNDNGQVFKINGSDTQGHVGAAGA 2049
 Db 2060 ASMPQAYIPGRSMVTGGTATHNGOGAVAGLSKLSNDNGQVFKINGSDTQGHVGAAGA 2119

Db 380 GLKEDSKKIVADVTALTVTGK-VAEIAKEDD-----KKLVNAG----- 419
QY 353 LKLVELANRKITPKGDDNN-----SNSVERGLGNTLTIKGDAQTNALTEANIGVVTD 406
Db 420 -DLVTALGNLSWAKABADTDGALEGISKQDEKAGETVTFKAG-----KNLKVQD 470
QY 407 GNLKVKLAKELTGLTSVATNKTIVTNTNNNAELQSGGLTFPIWCTKTDKTVYSIDG 466
Db 471 GANTYSLQDALTGLTS-----ITLGGTTN-----GGNDAKTIVNKDQ 508
QY 467 LKFTNDSNIAKGTTRITKKIKIFAGTNDGVDESKPVL-----DNEKLKVGNST---LN-- 518
Db 509 LTIIPAGNG-GTTGNTISVTKDKIKAGNKAITNVASCLRAYDDANEDVLNNSATDLNRH 567
QY 519 -----SGSLVNTNTGNKQIQVANGIKFATVANNVANTSATVGTARITEKIGF-AGTN 572
Db 568 VEDAYKGLLLNLNEKNANKPLV-----TDSATAVTGDLR-----KLGWVYSTK 610
QY 573 DGVDQAPYLDK--ERLKVGRVELITDGSINAGNHKITGLTNGTANT-----DAVTI 622
Db 611 NGTEESNQVQADDEVLTGAGAAVTVSKSNGKHITI---TVSVAETKADCGLEKDGDTI 667
QY 623 KOLDAKPTLNAGDGISINSNNGDLVDSSGNIITPTVYNISVKTTKLSNGTSGNKKESVS 682
Db 568 KLVVDNQNT-----DNVLTVGNNGTAVTKGGFET-----VKTGATDAD-----RCKVTVK 712
QY 683 NAHDNNS---LVTAKDLADLYLNKYNETADSALPSFKVQNGDNS---NNAITVVGKDTNGKT 736
Db 713 DATANDADKKVATVKDVATAIN-----SAATFVKTEINTLTSIDEDNPTDNGKDDALKA 765
QY 737 PNTLKLK-GEN-----GVNIT-----TNRATGTVTFGIDQSNGLT--TPKLTUVG 777
Db 766 GDTLTFKAGKNLKVCRDKGNTFLDAKLNLEKVTAKVSDTLTIGGNTPTGGTTATPKVNIT 825
QY 778 SDTNGNRLVIOVPASDGNSTKNI-IGLSPTL--BSTASPSGRNIALGNITIEEKDSNA 834
Db 826 STADGLNFAKE---TADAGSKNVLKGIATLTLEPSAGAKSSH---VDLNVDTATKKSNA 879
QY 835 ASIDLVLNAGNLKNNGKDKDFSVZTVDVFIDGNATATVTYDEANQTSKVAYDVNVDE 894
Db 880 ASIEDVLRAIGNIOGNGNVVYATVDVNFNTDSDTGTGTTVTVQ--KADCKGADVKIGA 937
QY 895 KTIELTDGNGKKQIKVTKTIKLTETSTNGNATFTFDDH-----ALVRAADI----- 941
Db 938 KTSVIKDHNGKLTG-KDLK-----DANNGATVSEDDGKDTGLTGLTAKTVIDAVNKS GW 991
QY 942 -----AGNLNTLAE--EIHHTTKGTANTALQTFVKKVDEN----- 974
Db 992 RVTGEGATAETGATVAVNAGNAETVTSCTSVPFKNGNATTA-----TVSKDNGNIINVKYDVN 1047
QY 975 -----DK--ADDTNAITVKGDTSGKY-----NTLKLKGKNGL----- 1005
Db 1048 VGDGLKIGDDKKIVADVTTLTV-----TGGKVSVPAGANSVNNKKLVNAEGLATALNLS 1103
QY 1006 -DIKTDKCTVTFGINTOSGLKAGDSTTLN-NGLSIKN-----TASNEQIQVGADGVKF 1058
Db 1104 WTAKADYADGESEGETDQEVKAGDKVTFKAGKNLKVQSKDEFTYSLQDTLTGLTSITL 1163
QY 1059 AMVNVVVGAGIDGTTRITREIGFTGNGS---LDKSKPH---LSKDGINAGKKIINI 1112
Db 1164 GGTANRNDTG-----TVINKDGLITLANGAAAGTDASNGNTISVTKDGISAGNKEITNV 1219
QY 1113 QSG-----EIAKNSHDVTGGKIDYDLKLELENKISSTAKTAQNSL 1152
Db 1220 KSALKYKDTQNTADETQDKFEHAAVKNANEVEFVGK-----NGATVSAKTDNNGK 1270
QY 1153 H-----EFSVADEQGNFTVSNPSSYDTSK-----TSDVI 1183
Db 1271 HTVTIDVAEKVGDLEKDKTDGKIKLKYDNTDGNLLTVDATKGAASVAKGEFNAVTTD 1330
QY 1184 TFAGENG-----ITTKVKNGVVRVIDQTKGLT----- 1211

Db 1331 TAQGTNANERKVVVKGSGNGATATETDKKKVATVCDVAKAINDAATFVKVENDSDSATIDD 1390
QY 1212 -----TPKLTVGNNG 1222
Db 1391 SPTDDGANDALKAGDTLTLAGKNLKVRRDGNITFALANDLSVKSATVSODKLSLGTNGN 1450
QY 1223 KGINVSONGON-----TITGLSNTLANVTNDKSGVBTTEQGNLIKDEDKTR 1269
Db 1451 KVNITSDFKGLNFRAKDSKTDGDANIHLNGIASTLTDTLNLSGAT-TNLGGNGITDNEKR 1509
QY 1270 AASIVDVLASGNLQG-----NGEAVDFSVTVDVTFNFGANTTTAKTVYD--DTSKT 1319
Db 1510 AASVKDVLNAGNVRGVKVPASANNQVENIDPVATVDTVDVFSGDKDTTTSVTVESKDNGR 1569
QY 1320 SKVYDVNVDDTTIEVKDKKLGKVTTLTSTG-----TGAN-KFALSNOAT-----G 1365
Db 1570 TEV-----KIGAKTSVIKDHNGKLTGKELKDDANNNGVTVTETDGKDEG 1613
QY 1366 DALYKASDIVAHLN'TLSGDIQTAKASQANNSAGYVD-----ADGNKVIYDSTDNKKY 1418
Db 1614 NGLVTAKAVIDAVNKAGRWVKTT-GANGONDDFATVASGTNVTFADGNGTTAEVT----- 1667
QY 1419 QAKNDG--TVDKTKEVAK-----DKLVAQA---QTPDGTLAQMNVKSVINKEQVNDANK 1467
Db 1668 -KANDGSTIVKYNVKVDGLKLDGKIVADTIVTLTVLTVADGKVTAPN-----NGDGK 1716
QY 1468 KOGINEDNAFYKLEAKASDMRKNAAVTVGLDNLNAVAOTPLTFFAGDTGTAKKLGELTIT 1527
Db 1717 K-----FVDSAGLADALNKLSWTA-TAGKEGTGEVDPANSAGQ-----EVKAGDKVTF 1763
QY 1528 KGGQDITNKLTDNNIGVVGAGTDGFTVKLAKOLTNLSV-----NAG-----GPKIDEKGIS 1578
Db 1764 KAG-----DNLKIKQSGKDEFTYSLKKELKDLTSEFEKDANGGTGSESTKITKDGLT 1814
QY 1579 FVDANGQA-----KANTPVLSSANGLDGKVSINSYKGTG-----DIDRAAN 1619
Db 1815 ITPANGACAAGANTANTISVTKDGISAGNKAVTNVSGLKKFGDCHTLANGTVADFEKHY 1874
QY 1620 VOQLNEVRNLLGLGNDN---ADGNQVNIAD-----IKKDPNSGS-----SSN 1658
Db 1875 DNAYKDLNLEKAGDNNPTVADNTAATVGLRLGLGWISADKTTGEPNOEYNAQVRNAN 1934
QY 1659 RTVVKAGT-----VLGG-----KKNNDTEKLTG---VOVG----- 1687
Db 1935 EVKFKSGNGIINVSGKTLNGTRVITFELAKEVYKVSNEFTVKNADGETNLVKVGDMYTSK 1994
QY 1688 VDKDGNANGDLSNVWVKTK-----DGSKKALLATVNAAGOTNYVTNNPAEATD 1736
Db 1995 EDIDPATSKPMTC---KTEKYKVENGVVANGSKTEVTLFNKSGS---YVTGN--QVAD 2046
QY 1737 RINEQIRFFHYNDGNQBPVVOGRNGIDSSASGRHSVAIGFOAK-----ADGEA 1785
Db 2047 AIAKSGFEL-----GLADAAEAFAESAKDKQLSKDKDAETVNAHDVK 2090
QY 1786 AVAIGROTOAGNOSTAIGDNAQATGDO-----SIAIGTGNVAGKHSAGIDPSTVK 1837
Db 2091 RFANGLNTKV---SAATVESTDANGDKVTTTFVKTVDLPLTQIY---NTDANGNRKIVK 2144
QY 1838 ADNSYSVGNNGQFTDATOTDVFVGNNITVTESNSVALGNSAISAGTHAGTQAKKSDGT 1897
Db 2145 ADGKW-----YELNADGTASNKEVTLGNVDANGKK-----VVKVTENG 2182
QY 1898 AGTTTATAGATGTVKGFAGQAVGAVSVG-----ASGABERRIONVAAGEVSATS 1945
Db 2183 ADKWTYNADGAADTKGEVSNDKYSTDEKHVVRVLDPNQNSNGKGVINDVANGEISATS 2242
QY 1946 TDVANGSOLY---KATQSTANATNLDHRIHONENKANAGISSAMAMASMPQAYIPGRSM 2002
Db 2243 TDAINGSOLYAVAKGVTNLAGOVNLEKGVKVRADAGTASALASQLPQATPMGRSM 2302
QY 2003 VTGGIATINGOGAVAGLSKLSNDNGQWVFKINGSADTOGHVGAAGVAGGFHF 2053
Db 2303 VAIAGSSVQGGQNGLAIGVSRISDNGKVIIRLSGTTNSOGKTGVAAGVGQW 2353

RESULT 4					
US-08-409-995-4					
Sequence 4, Application US/08409995					
Patent No. 5646259					
GENERAL INFORMATION:					
APPLICANT: Barenkamp, Stephen I.					
APPLICANT: St. Geme III, Joseph W.					
TITLE OF INVENTION: Haemophilus Adhesion Proteins					
NUMBER OF SEQUENCES: 6					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert					
STREET: Four Embarcadero Center, Suite 3400					
CITY: San Francisco					
STATE: CA					
COUNTRY: USA					
ZIP: 94111-4187					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					
SOFTWARE: Patentin Release #1.0, Version #1.30					
CURRENT APPLICATION DATA:					
APPLICATION NUMBER: US/08/409,995					
FILING DATE: 24-MAR-1995					
CLASSIFICATION:					
ATTORNEY/AGENT INFORMATION:					
NAME: Silva, Robin M.					
REGISTRATION NUMBER: 38,304					
REFERENCE/DOCKET NUMBER: A-61053/RFT					
TELECOMMUNICATION INFORMATION:					
TELEPHONE: (415) 781-1989					
TELEFAX: (415) 398-3249					
TELEX: 910 277299					
INFORMATION FOR SEQ ID NO.: 4:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 1912 amino acids					
TYPE: amino acid					
STRANDEDNESS: double					
TOPOLOGY: unknown					
US-08-409-995-4					
Query Match 8.6%; Score 890; DB 1; Length 1912;					
Best Local Similarity 23.8%; Pred. No. 3.5e-45;					
Matches 502; Conservative 227; Mismatches 680; Indels 702; Gaps 97;					
QY	1	MNHIYKVFENKATGTMVAEYAKSHISTGGSCATGVCS-----VRTLSFARIAALA--	53		
DB	1	MNKIFNVIMVMTQTWVVSELTTRTH-----TKLRNRGDPVLATLLFPATVOANATD	52		
QY	54	-----VLVIGA-----	59		
DB	53	EDELDPVRTAPVLSPHSDKEGTGEKVTENSNGIYFDNKGVLKAGAITLKAGDNLXX	112		
QY	60	-----TLNGSAYAAQI-----TTKIEIG-----	77		
DB	113	KQXTDXETNASFFYSKLKDLTDLTSVATEKLSFGGDKVDITSPANGKLAKTGNGNV	172		
QY	78	QTNKINTLKGDALATGEASIAFGSLSKAQSQAIAISGYKPPNNSGNSNGVSHAKGN-	136		
DB	173	HNLGDSTLPDAVNTGVLSSSFPPNDVEKTRA---ATVKDVLNAGWNIKGARTAGNV	229		
QY	137	ESIAATGGDVLAEGDASTAIGSD----DLYLPKNLDLKNFEHKLHGHLEIAKIQTSTDGI	193		
DB	230	ESVDL---VSAYNNVEFITGDKNTLDVVLTAKENXKTTEVRFPKPTSVIKE-----KDGL	282		
QY	194	---KYRRTRAQGHASTAVGAMSVAQGHFSNAFGTYATAEAAYSIAVLGAQAATQSSIAV	250		
DB	283	FYGKENDNTKVTSNTATDNTDEGNG-----LVTKAKAVID-AVNKAQRVRVTTTT---	330		
QY	251	GSAKANAFAPAATAIGGNTVNVNLGRGVALGFSGQILDRDNTDA-----SAYVPLG----	300		

Db 1276 VAEKVGDLGKEDTKIKLVNDTNDGNLLTVDATKCAVAKGEFNAVTTDAATTAQGTN 1335
Qy 1190 G-----ITTKVNGVVRVIDQTKGLT-----1211
Db 1336 ANERKGVVVGKSGATATETDKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDGG 1395
Qy 1212 -----TPKLTGVNNGKGIIVN 1228
Db 1396 ANDALKAXDTLTKAGKNLKYKRGKNTTFALANDLSVKSATVSDKLSLGTNGKNKNVITS 1455
Qy 1229 SONGON-----TITGSLNTLANVTNDKSGVTRTEQGNIKDEKDTAASTVD 1275
Db 1456 DTGLNFAKDSKTDGDANIHLNGIASTLTDILLNSGAT-TNLGGNGITDNEKKXAAASKVD 1514
Qy 1276 VLSAGFNLOG-----NGEAVDFSVYDVTNFANGNTTTAKVTYD--DTSKTSKVYVD 1325
Db 1515 VLNAGMNVGRVKPASANNQVENIDFVATYDVFVSGDKDTSVTVESKDNKGRTEV-- 1571
Qy 1326 VNVDTTIEVDKDKLGVKTTILTSTG-----TGAN-KFALSNOAT-----GDALVKA 1371
Db 1572 -----KIGAKTSVWIKDHNGKLTGCKELKDANNNGVTVTETDGRDEGNGLVTA 1618
Qy 1372 SDIVAHNLTSGLDIOTAKGASOANNSAGYVDADGNKVIYDSTDNKKYQAKNDGTVDKTKKE 1431
Db 1619 KAVIDAVNKAQRVKTT-GANGQNDFFATV-ASGTNVTF-----ADNGT---TAE 1664
Qy 1432 VAKDKLVAQAQTPDGTLL-AOMNVKSVINKEQVNDANKKQG---INEDNAFVKGLEKAASD 1487
Db 1665 VTK-----ANDGSITVKYVVK-----VADGLKLDGDKIIVADTVTLTVADGKVTAP 1709
Qy 1488 N-----KTKNAVTVGDLNVAQAQPLTFAGDTGT-----TAKKLGETLTIKGGQT 1532
Db 1710 NNGXGKFXDASGLAGCLNKLST--ATAGRECTGEVDPANSAGQEVKAGDKVTFKAG-- 1765
Qy 1533 DTNKLTDNNIGVAGTDCGFTVKLAKDLNLNSV-----NAG-----GTXIDKSGLSFVDAN 1583
Db 1766 -----DNLKIKQSKXKDFTSLSKELKDLTSVEFKDANGTGSESTKIKTKDGLTIITPAN 1818
Qy 1584 GOA-----KANTPVLISANGLDIGGKVISNGVGTGKTDDAANVQQLNEVRNLLGLGNDNAD 1638
Db 1819 GAGAAGANTANTISVTKDGISAGNKAIVNVVSLKK-----FGDGHHTLAN 1863
Qy 1639 GNOVNIADIKK 1649
Db 1864 G---TVADFEK 1871

RESULT 5

US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELECOMMUNICATION INFORMATION:
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-685-467-4

Query Match 8.6%; Score 890; DB 3; Length 1912;
Best Local Similarity 23.8%; Pred. No. 3.5e-45;
Matches 502; Conservative 227; Mismatches 680; Indels 702; Gaps 97;

Qy 1 MNHIYKIPNKGATGFMVAEYAKVSHSTGGSCATGQVGS-----VRTLSFARIAALA-- 53
Db 1 MNKIPNVIWNVMTQWVWVSELTRTH-----TKRLNRNGDPVLATLLFPATVQANATD 52
Qy 54 -----VLVIGA----- 59
Db 53 EDEELDPVVRAPVLSFHSKDEKTEKEVTENNNNGIYFDNKGVLKAGAITLAKGDNLXX 112
Qy 60 -----TLNGSAYAOOI-----TTKTEIG----- 77
Db 113 KOXTDEXTNASSFTYSLAKDLTDLTSVATEKLSFGANGDKVDITSDANGLKLAKTNGNV 172
Qy 78 QTNKINNTLKGDALATGEASTAFGSLSKAQSQAIAIGSVKPPDPNNGSNGVSHAKGN- 136
Db 173 HNLGDLSTLPOAVTNTGVLSSSFTPDNVEKTRA---ATVKDVLNAGNIIKAGAKTAGNV 229
Qy 137 ESIAIGGDVLAEGDASIAIGSD--DLVLPKNLDLKNFEHKLHGHETLKKLTOTSDGKI 193
Db 230 ESDVL---VSAYNVVEFITGDKNTLDVLTAKENAKTTEVRETPKTSVIKE-----KDGKL 282
Qy 194 ---KYRTRRAOCHASTAVGAMSIAOQHFENAFGTATAEAAVSLAVGLAAQATKOSSTAV 250
Db 283 FTGKENNDTKVTSNTATDNTDEGNG-----LVTAKAVID-AVNKAGWRVKTT--- 330
Qy 251 GSNKANAFAATAIGNTVNVNLGRVGLFGSQIILDRDNTDA-----SAYVPLG----- 300
Db 331 -ANGONGDFATVASGNTVTFESGDGTTAS-----VTKDTNGNGITVYDAKVGDLKLFDS 384
Qy 301 --KTLADQYKATROGSDTDIFSIGNNNNNSSIRKIIINVAGSROTDVAVNVAQLKLVEE 358
Db 385 DKKIVADTTALTVTGKG--VAEIAKEDD-----KKLVNAG-----DLVTA 423
Qy 359 LANRKITEFGDGDNN-----SNSVERGLGNLTITKDAQTNALTEANIGVVVTDGNGLVK 412
Db 424 LCNLSWKAKAEADTGALEGISKDEKAGEVTFKAG-----KNLVKQDQGANFTY 475
Qy 413 KLAKELTGLTSVATNKITVSNNTNNNAELOSGLTFFSPITGKTGKTIVYSIDGLKFTND 472
Db 476 SLQDALTLGLTS-----ITLGGTTN-----GNDAKTVINKDGLTITPA 513
Qy 473 SNSIATKGTTRITKKIKGFAGTNDGVDESKPYL---DNEKLVGNST---LN-----S 519
Db 514 GNG-GTTGTNTISVTGDKIGAKNKAITNVASGLRAYDDANFDVNNNSATDLNRHVEDAYK 572
Qy 520 GSLTVNNTTGNKQIOVGANGTKFATVANNVANTSATVGTARITEKIGF-AGTNDGVDEO 578
Db 573 GLLNLNEKNANKQPLV-----TDSATATVGDLR-----KLGWVYSTKNGTKEE 615
Qy 579 APYLDK--ERLKVGRVEITDTSGINAGNHKITGLTNGIANT-----DAVTIKQLKDA 628

Db 616 SNQVQADEVLEFGAGAAVTSKSENGKHTI---TVSVAETKADCGLEKDGDTIKLVND 672
 QY 629 KPTLNAGDGLSINSNGDLVDSSGNITPTPIYNSVKTTLKNSNGTSGNNKFSVSNHADDN 688
 Db 673 QNT---DNVLTGVNGNGTAVTKGFFET-----VKTGATDAD-----RGKVTVKDATAND 717
 QY 689 S---LVTAOLADYLNKVNETAADALPSFKVQNGDNS---NNAITVGKDTNGKTFNTKL 742
 Db 718 ADKXVATVQVATIN-----SAATFVKTENLTYSIDEDNPTDNGKODALKAGDTLIF 770
 QY 743 K-GEN-----GNVIT-----TNRAITGVTEGIDOSNGLT---TPKLTVGSSTNGN 783
 Db 771 KAGKLNKVKRGKGNITPDALKNLEVTAKVSDTLTIGGNTPTGTTATPKVNITSTADGL 830
 QY 784 RLVEIQVPSADGNSTKNI--DKGLSPTL--PSIASPSGRNIALGNTTEKDKSNAASIDV 840
 Db 831 NFAKE---TADAGSKNYYLKGIAITLITTEPSAGAKASH---VDLNDVATKSNASIEDV 884
 QY 841 LNAGFNLNKNGKDKDFVSTYDVTDFIDGNATTATVYDEANQTSKAYDVYNDVDEKTIET 900
 Db 885 LRAGWNIQGNVNDVYVATVDTVNFTDSTGTTVTVTQ--KADGKADVKIGAKTSVK 942
 QY 901 GDNGKKOLGKTKLTSTSTNGNATFTSTDDH-----ALVKAADI----- 941
 Db 943 DHNGKLEFG-KDLK-----DANNGATVSEDDGKDTGTLVTAKTVIDAVNKGSRVTEG 996
 QY 942 -----AGNLNTLAE--EIHHTKGTANTALQFTVKKVDEN----- 974
 Db 997 ATAETGATVAGNAEIVTSGTSVFNFGNATTA---TVSKDNGNINVKYDVYNGVGDGLK 1052
 QY 975 ---DK---ADDTNATVYKDGDTSGKV-----NTLKLKGNGL-----DKTD 1010
 Db 1053 IGDDKKIVADTTTLTV---TGGKVSVPAGANSVNNKLVNABEALATLNNLSWTAKAD 1108
 QY 1011 KDGTVTEGINTQSLKAGDSTTLN-NGNLSKN-----TASNEQIOVGADGVKFAVYNG 1064
 Db 1109 KYADGESEGTQVQKAGDKVTFKAGKLNKVKOSEKDTYSLQDTLGLTSITLGGTANG 1168
 QY 1065 VVAGIDGTTTRITDEIGFTNGS---LQSKPH---LSKRGINAGKKTITNQSG--- 1115
 Db 1169 RNDTG---TVNKDGLTITLANGAAAGTDAANGNTISVTKDGLSAGKEITVNVKSALKT 1224
 QY 1116 -----EIAKNSHDVATGKIYDLKTELENKISSAKTAKNSLH----- 1153
 Db 1225 YKDTQNTADETQDREFAAUVKANEVEFVK-----NGATVSAKTDNNGKHTVTD 1275
 QY 1154 -----EFSVADQGNFTVSNPYSSYDTSK-----TSDVITFAGEN 1189
 Db 1276 VAEKVGDLKEDTDGKIKLVNDTGNLLTVDATKGASVAKGEFNAVTTDATAGTN 1335
 QY 1190 G-----ITTKVYKGVVRVIGIDQTKGLT-----TPKLTGVNNGKGVIN 1211
 Db 1336 ANERKVVVKGSGATATETDKKVVATGVDAKAINDAATFVKVENDDSATIDDSPTDDG 1395
 QY 1212 -----TPKLTGVNNGKGVIN----- 1228
 Db 1396 ANDALKAXDTLTLKAGKNLVKVRGKNTFPALANDLSVKSATVSDKLSLGTNGKVNITS 1455
 QY 1229 SONGON-----TITGLSNTLANVNDKGSVRRTEQGNIIKDEKTRAASIVD 1275
 Db 1456 DTKGLNFAKDSKTGDANIHLNGIASTLTDLLNSGAT-TNLGGNGITDNEKKXAAASKVD 1514
 QY 1276 VLSAGFNLOG-----NGEAVDFVSTYDVNFANGNTTAKVTYD--DFSKTSKVYVD 1325
 Db 1515 VLNAGWNVRGVSPASANNQVENIDFVATYDVFVSGDKDTSVTVESKONGKRTVE--- 1571
 QY 1326 VNVDDTIEVKDKLGKVVTTLTSTG-----TGAN-KFALSNOAT-----GDALVKA 1371
 Db 1572 -----KICATSVIKDHNGKLFVTKELKDNANGNVITVETDGDGEGNGLVTA 1618
 QY 1372 SDIVAHNLNLSGDIQTAKGASOANNSAGYVDADGNKVTYDSTDNKYQAKNDGTVDKTK 1431

Db 1619 KAVIDAVNKGARVKT--GANGQNDQFATV-ASGTNVT-----ADNGT---TAE 1664
 QY 1432 VAKDKLVAQAQPDGTL-AQMNKVSINKQVNDANKKQ-----INEDNAFVKGLEKAAASD 1487
 Db 1665 VTK-----ANDGSIIVKYNV-----VADGLKLDGDKIVADTTVLTVADGKVTAP 1709
 QY 1488 N-----KTNNAVTVGDLNVAQAQTLTFAGDTGT-----TAKKLGTELTLTKGGOT 1532
 Db 1710 NNGXGKKFADAGSLAGCLNKLST--ATAGKEGTGEVDPANSAGQEVKAGDKVTFKAG-- 1765
 QY 1533 DTNKLTDNNIGVVGAGTGTGFTVKLAKDLTNLSV-----NAG-----GTKIDEXGISFVDAN 1583
 Db 1766 -----DNLKIQSKKDTYSKLKELKDLTSVEFKDANGGTGSESTKITKGLTITPAN 1818
 QY 1584 GOA-----KANTPVLVSANGLDLGGKVISVNGKGTKPTDAANYOOLNEVNNLGLGNDNAD 1638
 Db 1819 GAGAAGANTANTISVTKDGISAGNKAFTNVVSGLKK-----FCDGHTLAN 1863
 QY 1639 GNOVNIADIKK 1649
 Db 1864 G---TVADFEK 1871

RESULT 6

US-08-409-995-2
 ; Sequence 2, Application US/08409995
 ; Patent No. 5646259
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen I.
 ; APPLICANT: St. Gene III, Joseph W.
 ; TITLE OF INVENTION: Haemophilus Adhesion Proteins
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/409,995
 ; FILING DATE: 24-MAR-1995
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Silva, Robin M.
 ; REGISTRATION NUMBER: 38,304
 ; REFERENCE/DOCKET NUMBER: A-61053/REF
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1098 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: unknown
 ; US-08-409-995-2

Query Match 6.7%; Score 690; DB 1; Length 1098;
 Best Local Similarity 24.7%; Pred. No. 1.6e-33;
 Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;
 QY 953 HTTKGTANTA---LQTFVKKVVDENKADDTNATVTKDGTSGKVT----- 996
 Db 26 HTRCASATVAVLATLLSATVEANNPTVTKLKAYGDANFNFTNNSIADAKEQVQEAY 85

QY 997 ---LKLKKNGLDIKTDKGTIVTGINTQSLKAGSDSTLNNNG--LSIKNTASNEQIQV 1051
Db 86 KGLLNLEKKNASDKLLVEDNT-----AATVGNLURKLGWLSSKNGTNRKESQ- 132
QY 1052 GADGVKFAVNVNNGVVGAGIDGTTTRITRDEIGTGTNGSLDKSPHLSKDGINAGCKKITN 1111
Db 133 ---QVQKHA-----DEVLFEG-----KGVQVTS 152
QY 1112 IOSGEIAKNSHDAVVGKIKYDLKTELENKISSATAQNSLH--EFSVADQGNFTVSN 1169
Db 153 -----TSENGKHITITFALAKDLG----- 170
QY 1170 PYSSYDTSKSDVITFAG--ENGITTT--KYNKGVVVRGIDQTKGLTTPKLTGVNNGKGI 1225
Db 171 ---VKTATVSDTLTIGGAAAGATTPPKVNV-----TSTTDGLKFAKDAAGANGDPTV 220
QY 1226 VINSQNGONTITGLSNTLANVTNDKSGVRTEQNIILKDEK---TRAASIVDLSAGFNL 1283
Db 221 HLN-----GIGSILTDPL--VGSPTHIDG---DQSTHYTRAASIKDVLNAGWNI 266
QY 1284 Q-----GNGEAVDFSVYDVFANFGNTTTAKVTYDDTSKSVVYDVNVDDTIEV 1335
Db 267 KGVKAGSTGGOSENVDFVHTYDTEFLSADPTTTVTVDSEKNGKRTKVKIGAKTSVIKE 326
QY 1336 KDKKLGK-KTTTLTSTGTGANKFALSQAOTGDALVKASDIVAHLNLTSGDIQTAKGASQA 1394
Db 327 KDKGLFTGKANKEFKVKGAN--ATEDADEGKGLVTAQDVDAVNTKTCWRIKTTDANGON 384
QY 1395 NNSAGYVD-----ADGN---KVYDSTDNKYQAQKNDGTVDKTEKAKDKLVAQAQTPD 1445
Db 385 GDFATVASGNTVTFASNGTTATVNTGTDG--ITVYDAKVGDLGLDKGXIAA-----D 437
QY 1446 GTLAQMNVKSVINKEQVNDANKKGINEEDNAFVKGLEKKAASDNK---TKNAAVTVGDINA 1502
Db 438 TTALTVN-----DGKNANPKGVADVASTDEKKLVTAKGLVTA--LNS 479
QY 1503 VAQTPPLTFAGDTGT-----TAKLGETLTIKGGQDITNKLTDNNIGVYAGTGDGFTVKL 1555
Db 480 LSWTTTAAEADGGTLDGNASEQEVKAGDKVTFKAG-----NLKVQEGGANFTYSL 530
QY 1556 AKDLTNLSV-----NAGTKKIDEKGISFVDANGQA--KANTPVLSSANGLDLGGKVISN 1607
Db 531 QDALTLGLTSITLGTGNGCAKEINIKDGLTTPANGAGANNANTISVTKDGLISAGGQSVKN 590
QY 1608 VKGTKDITDAANVOOLNEVRNLGLGNDNAGQVNTADIKKDPNSGSSSRNRTVIKAGTV 1667
Db 591 VVSGLKFKFGDANFDPDITSSADNLTKQNDAYKGLTNLDEKGTDKQTPVVAADTAATVGD- 649
QY 1668 LGKGNNDTEKLATGGVQGVYDKDGNANGDLSNVVKTQDKSKKALLATYNAAGQTYV 1727
Db 650 LRLGLWISADKTTGGSTYHDQVRNANE-----VKF-KSGN-----GINVSGKT--- 693
QY 1728 TNNPAEATRINEQIRFFHFVNDG-----NQBPVVQGRNGIDSSASGKHSVAIG--FOAKA 1781
Db 694 -----VNGRREITFELAKEVEVKSNEFTVKEINGKETSLL-----VKVGDKYYSKE 738
QY 1782 D-----GEAAVAIGRQTOA-----GNOSIATGDNAQATGDOOSIAIGTGNVYAGKHSAGIG 1831
Db 739 DIDLTTGQPKLKDGTNVAAYKODRGKGVSVTDNTEATITNK-----GSGYVTGNQVADAI- 794
QY 1832 DPSTVKADNSYSVGNNO-----FTDATQDVFVGNNITVTSNVALGSNAISAGT 1885
Db 795 -----AKSGPELGLADEADAKRAFDDKTKALSAGTTEIVNAHDKVRFANGLNTKVSAA 848
QY 1886 HAGTQAK-----KSDGTACTT 1901
Db 849 VESTDANGDKVTTTFVKTDVPLQIYNTDANGKKITKVVYKDGQTKWYELNADGTADMT 908
QY 1902 TT-----AGATGTVKGFAGQATVAGVSVG-----A 1926
Db 909 KEVTLGNVDSGKVKVNDKDGKWKYHAKADGTADTKTKEVSNDKYSTDEKHVVSLDPNDQS 968
QY 1927 SGAERRQNVAAGEVSATSTDAVNGSOLY---KATQSIANAATNELDRIHQENKANAGI. 1983

Db 969 KGKGVVINDVANGDISATSTDAINGSQLYAVAKGVTNLAGOVNNLLEGKVNKVGKRADAGT 1028
QY 1984 SSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAGLSKLSKSDNGOWFKINGSADTQGHV 2043
Db 1029 ASALAASOLPOATWPKSKVAIAGSSYOGONGLAIGVSRISDNGKVIIRLSGTTNSQSKT 1088
QY 2044 GAAYGAGPHF 2053
Db 1089 GVAAGVGYOW 1098
RESULT 7
US-08-685-467-2
; Sequence 2, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M. 38,304
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-685-467-2
Query Match 6.7%; Score 690; DB 3; Length 1098;
Best Local Similarity 24.7%; Pred. No. 1.6e-33;
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;
QY 953 HTTKGTANTA---LQTFVTKVVDENDKADDTNATVTKDGTSGKVT----- 996
Db 26 HTKASATVAVAVLATLTLSATVEANNPTVNTKLYGDANFNFTNNSIADAEKQVQRAY 85
QY 997 ---LKLKKNGLDKTDKDGTVTGINTQSLKAGSDSTLNNNG--LSIKNTASNEQIQV 1051
Db 86 KGLLNLEKKNASDKLLVEDNT-----AATVGNLURKLGWLSSKNGTNRKESQ- 132
QY 1052 GADGVKFAVNVNNGVVGAGIDGTTTRITRDEIGTGTNGSLDKSPHLSKDGINAGCKKITN 1111
Db 133 ---QVQKHA-----DEVLFEG-----KGVQVTS 152

Thu Sep 13 14:19:21 2001

QY 1112 IOSGEIAKNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLH--EFSVADQGNFTVSN 1169
Db 153 -----TSENGKHTITFALAKDLG----- 170
QY 1170 PYSSYDTSKTSVITFAG--ENGIT--KVNGVVRVGIDQTKGLTTPKLVGNNGKGI 1225
Db 171 ----VKTATVSDTLTIGGGAAGATTTPKVV-----TSTTDGLKFAKDAAGANGDTTV 220
QY 1226 VINSQONQITGLSNFLANVTNDKGSVRTTEQNIIDEDK--TRAASIVDVLGAGFNL 1283
Db 221 HLN-----GIGSLTDLT--VGSPTHIDGG--DQSTHYTRAASIKDVLNAGWNI 266
QY 1284 Q-----GNGRAVDVSVTYDVFNFANGNTTAKVYDDTSKTSKVYDVNVDDTTIEV 1335
Db 267 KGVKAGSTTGOSENVDFVHYDVFELSDATETTTVTVDSEKNGKRTVEKIGAKTSVIRE 326
QY 1336 KDKKLG--KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIOTAKGASQA 1394
Db 327 KDKLFTGKANKETNKVDGAN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGON 384
QY 1395 NNSAGYVD-----ADGN---KVIYDSTDNKYQAKNDGTVDKTEKAKDLVAQAQTPD 1445
Db 385 GDFATVAGSTNVTTFASGNGTTATVTNGTDG--ITVKYDAKVGDLKLDGDKIAA-----D 437
QY 1446 GTLAQMNVKSVINKEQVNDANKKOGIENEDNAFVKGLEKAASDNK---TKNAAVTVGDLNA 1502
Db 438 TTALTIVN-----DGKANNPKGVADVASTDEKKLVTAAGLVTAA--LNS 479
QY 1503 VAQTPLETFAGDTGT-----TAKKLGTLTIKGGQDTNKLTDNNIGVVAGDTGGTFVKL 1555
Db 480 LSWTTTAAEADGTLUDGNASQEVKAGDKVTFKAG-----NLKVKQEGANFTYSL 530
QY 1556 AKDLNLNSV-----NAGGTIDEKGLISFVDANGQA--KANTPVLNANGLDLGGKVISN 1607
Db 531 QDALTLGLTSITLGTGNGAKTEINKDGLTITPANGAGANNANTLSVTKDGLSAGGQSVKN 590
QY 1608 VQKGTQDAAVQOLNEVRNLLGNDNADGNQVNIADIKDPNCSGSSSRNRTVIKAGTV 1667
Db 591 VVSGLKKFGDANFDPLTSSADNLTKQNDAYKGLNLDEKDTKQTPVADNTAATVGD- 649
QY 1668 LGKGHNDDTEKATGQVQGVQDKGANGNDLSNVWVKQDKGSKALLATYNAAGQTNV 1727
Db 650 LRLGLWVISADTKTGGSTHYHQVRNANE-----VKF-KSGN-----GIVNSGKT--- 693
QY 1728 TNPAEALDRINQOIRFEHVNDG-----NOEPVQGRNGIDSSASGKHSVAIG--FOAKA 1781
Db 694 -----VNGRREITFELAKGEVVKVNSFEFTVRETNGKETS-----VKVGDKYYSKE 738
QY 1782 D-----GEAAVATGROTA-----GNQSTAGDNAQATGDQSIATGTVNAGKHSAGIG 1831
Db 739 DIDLTTGPKLKDGTVAARYQDKGVVSVTDNTEATITNK-----GSGVTVGNQVADAI- 794
QY 1832 DPSTVKADNSYSVGNQ-----FTDATQTDVFGVGNNTVTESNSVALGSNSAISAGT 1885
Db 795 -----AKSGFELGLADEADAKRAFDDKTKALSAGTTEIVNAHDKVRFANGLNTRKVSAA 848
QY 1886 HAGTQAK-----KSDGTAGTT 1901
Db 849 VESTDANGDKVTTTFVKTVDVLPITQIYNTDANGKTKTKVVDGQTKWYELNADGTADMT 908
QY 1902 TT-----AGATGTVKVGFAGQTAAGVAVSG-----A 1926
Db 909 KEVTLGNVDSGKKVVDNDNGKWYHAKADGTADTKTKEVSNDRKVSDEKHVVSLDPNDQS 968
QY 1927 SGAERRIONAAGEVSATSTDAVNGSOLY---KATQSIANATNELDRHONENKANAGI 1983
Db 969 KKGKGVVDNVAINGDISATSTDAINGSQLYAVAKGVTVNLAGOVNLEKGVNKKVCKRADAGT 1028
QY 1984 SSAMAMASMPQAYIPGRSMWTTGATATNGQAVAGVLSKLSDNQGWYFKINGSDAQGHV 2043
Db 1029 ASALAAASQLPQATMPCKSNVAIAGSSYOGQGLAIGVSRISDNKGVIIIRLSGTTNSQOKT 1088

QY 2044 GAAVGAGPHF 2053
Db 1089 GVAAGVGYOW 1098
RESULT 8
US-09-377-155-32
; Sequence 32, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRK
; ORGANISM: Haemophilus influenzae
US-09-377-155-32
Query Match 6.7%; Score 690; DB 4; Length 1098;
Best Local Similarity 24.7%; Pred. No. 1.6e-33;
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;
QY 953 HTTKGTANTA---LOTFTVKVVDENDKADDTNATTVGKDTSGKVT----- 996
Db 26 HTKCSATVAVAVLATLSATVEANNPTPVINKLKAYCDANFNFTNNSIADAEKQVQEA 85
QY 997 ---LKLKNGKGLDIDKTDGVTGINTQSGKAGDSTTLNNG--LSIKTASNEQIOV 1051
Db 86 KGLLNLEKNASDKLLVEDNT-----AATVGNLRLKGLWVLSRRKNGTRKESQ- 132
QY 1052 GADGVKFAVNVGVGAGIDGTTRITRDEIGFTGNGSLDKSKPHLSKDGINAGGKITT 1111
Db 133 ---QVKAH-----DEVLEF-----KGVQVQTS 152
QY 1112 IOSGEIAKNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLH--EFSVADQGNFTVSN 1169
Db 153 -----TSENGKHTITFALAKDLG----- 170
QY 1170 PYSSYDTSKTSVITFAG--ENGIT--KVNGVVRVGIDQTKGLTTPKLVGNNGKGI 1225
Db 171 ----VKTATVSDTLTIGGGAAGATTTPKVV-----TSTTDGLKFAKDAAGANGDTTV 220
QY 1226 VINSQONQITGLSNFLANVTNDKGSVRTTEQNIIDEDK--TRAASIVDVLGAGFNL 1283
Db 221 HLN-----GIGSLTDLT--VGSPTHIDGG--DQSTHYTRAASIKDVLNAGWNI 266
QY 1284 Q-----GNGRAVDVSVTYDVFNFANGNTTAKVYDDTSKTSKVYDVNVDDTTIEV 1335
Db 267 KGVKAGSTTGOSENVDFVHYDVFELSDATETTTVTVDSEKNGKRTVEKIGAKTSVIRE 326
QY 1336 KDKKLG--KTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIOTAKGASQA 1394
Db 327 KDKLFTGKANKETNKVDGAN--ATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGON 384
QY 1395 NNSAGYVD-----ADGN---KVIYDSTDNKYQAKNDGTVDKTEKAKDLVAQAQTPD 1445
Db 385 GDFATVAGSTNVTTFASGNGTTATVTNGTDG--ITVKYDAKVGDLKLDGDKIAA-----D 437
QY 1446 GTLAQMNVKSVINKEQVNDANKKOGIENEDNAFVKGLEKAASDNK---TKNAAVTVGDLNA 1502
Db 438 TTALTIVN-----DGKANNPKGVADVASTDEKKLVTAAGLVTAA--LNS 479

FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA: PCT/US96/4031
APPLICATION NUMBER: 22-MAR-1996
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RET/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 679 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-913-942-15

Query Match 5.6%; Score 583.5; DB 4; Length 679;
Best Local Similarity 23.9%; Pred. No. 2.1e-27;
Matches 226; Conservative 115; Mismatches 223; Indels 383; Gaps 34;

Qy 1135 TELEN-KISSTAKTAQNSLHEFSVAD-----EQNNFTVS-----NPYSSYDTSKTSKSDVI 1183
Db 88 TEVINLNTDSSGNVGSSTITFKAGDNLKIKQSGNDEFTYSLKKELKNTLSVETKLS--- 144

Qy 1184 TFAGENGITTKVNGVVRVGGIDQTKLTPKLT-VGNNGKGVINSQNGQNTITGLSNT 1242
Db 145 --FGANG-----NK-----VDITSDANGLAKLTGNGG-----QNSNVHLNGIAST 184

Qy 1243 LANVT--NDKGSVRTTEOGNIKDEKTPRAASIVDVLSAGENLOGNEAVDFVSTYDTVN 1300
Db 185 LDTLGGTTGHVDN-----IDAVNYHRAASVQDVLNNGNIOGNNVDFVRYTIDVD 239

Qy 1301 FANGNTTTAKVTYDDTSKTSKVYDVVDYDDTTIEVKDKKLGKVTKTTLTSTGTGANKFALS 1360
Db 240 FVNGANANVSVDTAHAKT---TVRVDVGLPVQ----- 271

Qy 1361 NOATGDALYKASDIVAHLNLTSGDIQTAKAGSQANNSAGYVDADGNKVIYDSTONKYYQA 1420
Db 272 -----KTK----- 311

Qy 1421 KNDGTVDKTKEVAKDKLVAQAOTPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFVKG 1480
Db 291 KDDGSADMMQKV-----ENGELA----- 308

Qy 1481 LEKAASDNKTKNAAVTVGDNLNAVAQTPPLTFAGDTGTAKLGLSETLTIKGGQDITNKLTDN 1540
Db 309 -----KTK----- 311

Qy 1541 NIGVAGTDGFTVKLAKDLTNLSNVNAGGTDKIDEGISFVDANGQAKANTPVLISANGLDL 1600
Db 312 -----VKL-----VSAGT-----NPV----- 323

Qy 1601 GGVISNVGKGTGDTDAANVQOLNEVRNLLGLNDNADGNQVNIADIKDPNSGSSSNT 1660
Db 324 ---KISNVADGTDGDAVSEFKALQD-----KQVTLSTSNAYANGGTDND--- 367

Qy 1661 VIKAGTVLGGKGNNDTEKLTATGGVQGVVD-KDGNANGDLSNVVWVKTKQDGSKKALLATYN 1719
Db 368 -----GGK-----ATQTLNGLNFKFKSSDGLLKI-----S 394

Qy 1720 AAGTNTVNNPAAIDRINEQGRIFHFVNDGNOEPVVGVRNGIDSSASGKSHSVATGFQA 1779
Db 395 ATGDT--VTFTPK----- 412

Qy 1780 KADGEAAVAIGRQTOAGNOSTAIGDNAQATGDSIAIGTGNVVGKHSKA-----IGD 1832
Db 413 -DDGKASISKAGNTEG--LVEASELVESLNKLGWKVGVEKVGSGELDGTSKETLVKSGD 469

Qy 1556 AKDLTNLSV-----NAGGTXIDEKISFVDANGQA--KANTPVLISANGLDLGGKVISN 1607
Db 531 QDALTGTLSTLTGTGNGAKTEINKDGLTITPANGAGANNANTISVTKDGISAGQSQSVKN 590

Qy 1608 VGGTGTDTDAANVQOLNEVRNLLGLNDNADGNQVNIADIKDPNSGSSSNTVTKAGTV 1667
Db 591 VVSLGKKFGDANFDPLTSSADNLTKQNDAYKGLTNLDEKGTGDKQTPPVADNATATVGD- 649

Qy 1668 LGKGKNNDEKLTATGGVQGVVDKGNANGDLSNVVWVKTKQDGSKKALLATYNAAQOTNVV 1727
Db 650 LRLGLWVISADKTTGGSTEHDQVRNANE-----VKF-KSGN-----GINVSKT--- 693

Qy 1728 TNPAEAIDRINEQGRIFHFVNDG-----NOEPVVGVRNGIDSSASGKSHVAIG--FQAKA 1781
Db 694 -----VNGRREITPELAKGEVVKVNEFTVKETNGKETS-----VKVGDRYYSKE 738

Qy 1782 D-----GEAAVAIGRTOQA-----GNQSTAIGDNAQATGDSIAIGTGNVVGKHSKAIG 1831
Db 739 DIDLTGQPKLKDGTNTVAARYQDKGVKVSVDTEATITNK---GSGYVTGNQVADAI- 794

Qy 1832 DPSTVKAADNSVSGNNNO-----FTDATQTDVFGVGNNTITVESNSVALGNSAISAGT 1885
Db 795 -----AKSGFELGLADEADAKRAFDKTKALSAGTTEIVNAHDKVRFANGLNTKVSAAAT 848

Qy 1886 HAGTOAK-----KSDGTAGTT 1901
Db 849 VESTDANGDKVTTTFVKTVDVELPLQIYNTDANGKKTIVKVKDQGTQKVELNADGTADMT 908

Qy 1902 TT-----AGATGTVKGFAGQTAGVAVSVG-----A 1926
Db 909 KEVLGNVDSGKKVVKDNDGKWTYHAKADGTADKTKGEVNSDKVSTDEKHVWSLDPNDQS 968

Qy 1927 SGAERRIONAAGEVSATSTDAVNGSQLY---KATQSIANATNELDHRHONENKANAGI 1983
Db 969 KKGKVVIVDANGDISATSTDAINGSOLYAVAKGVTNLAGOVNLEKGVNKNVGRADAGT 1028

Qy 1984 SSAMAKASMPQAYIPGRSMVTTGGTATHNGQAVAGLSKLSLDNGQWFKINGKISADTQGHV 2043
Db 1029 ASALAASQLPQMTPEKSWAIGSSYQOGNGLAIGVSRISDNGKVIIRLSGTTNSOCT 1088

Qy 2044 GAAVGAGHFH 2053
Db 1089 GVAAGVGYQW 1098

RESULT 10
US-08-913-942-15
Sequence 15, Application US/08913942
Patent No. 6200578
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehr Hohnbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995

Db 443 NVTIAPSASRVELGADRNH--SAEVIKVTLKNNITSLTTLTNTT-----ISNLIKS 493
QY 543 ATVANNVANTSATVGTARITEEKIFAGTNDGVDEQAPYLDKERLKVGRVEITDSDG---599
Db 494 AHVVNITARRKLTWNSSISIERGSHLLHSEGGOGGQVQIDK-----DITSEGGNLT 545
QY 600 INAGN-----HKITGLTNGIANTDAVTIKQ-----LKD-----AKPTLNAG-----635
Db 546 IYSGGWVDVHKNTITLGSGLN---ITTKEGDIAFEDKSGRNNLTITAGTITSGNSNGFR 602
QY 636 -DGISINSNGDL--VDS-----SGNLT-----TPVTNISVKTTKL-----669
Db 603 FNNVSLNSLGGKLSFTDSREDRGRRTKGNISNKFDPDGLTNSIGTVDISNKAAPKVSFWYRDK 662
QY 670 -----SNGTSGNNKFSVS-----NAHDNNSLVTAOKLADLYLNKVNETAADSALPSFK 715
Db 663 GRTYWNVTTLNVTSG-SKFNLSIDSTGSGTGPSIRNAE-----LNGI--TFNKA--TFN 712
QY 716 VONGDNNNAITVG-----KDTNCKTEN-TLKLKGENGVNITTNRATCTVFGIDQSNGL 769
Db 713 IAQGSTANFISIKASIMPFKSNANYALFENEDISVSGGVSFNKLNAS-----SSNI 762
QY 770 TTPKLTVCSDTNGNRLVIEQVPSADGNSTKNI-IKGLSPTLPSIASPSGRNIALGN-TIE 827
Db 763 QTPGVLIKS-----QNFVWSGGSTLNKKAEGTETAFSIENDLNUNATGNTITIR 812
QY 828 EKDSNAASIDDLVNLNAGFNKNGKDRDFYSTVDYDFIDGNATTATVYDEANQTSKVA 887
Db 813 QVE-----GTDSDRVNKGVAACKN-----ITFKGN-----ITGSKATTEIK 850
QY 888 YDWNVDEKT-IELTGDN---GKKQLGVKTKLKTETSTNGNATFTSTDDHALVKASDIAG 943
Db 851 GNVITINKNTATNLKAGFAENKSPNLT---AGNVINNGNLT-----AG 891
QY 944 NLNPLAEIHTTKTANTALQTFVKKVDENDKADDTNATVKGDTSGKVNTLLKAGN 1003
Db 892 SIINIAGNLVSKANLOAITNYIFNVAGSEFDDNGASN-LSIARGGAKFK-----DINNYS 946
QY 1004 GLDIKTDGDTVTFTGINTQSLGKAGDSTTLNNGNLSIKNTASNBQIQVADGVKFAVMNN 1063
Db 947 SLNITNSDTPYRITKGNISNKGSD-----LNIIDKKSDAEIQIGN-----989
QY 1064 CVVAGIDGCTTRIDREIDGFTGCTGSLDKSKPHLSKOGIAGGKKIYNIOGSETAKNSHD 1123
Db 990 -----ISQKEGNTLISDKYNTINQITIKRAGVEGSDSSEAEANLTIQTKE 1037
QY 1124 AVTGGKIYDLKTELENKISSAKTAQNSLHEFSVADQGNFTVSNPYSSVDTSKTSQV 1183
Db 1038 LKLAG---DLNISGFNAEITAKNGS-----DLTIGNASGGNADAK--VTFDKVKDSKIS 1088
QY 1184 TFAGEN---GITTQVKNKGVVRVIGDQTKGLT--TPKLTVGNN-----N 1221
Db 1089 T-DGHNVTLNSEVKTSNGSSNAGNDSTGLTISAKDVTNNVNTSHKNTINISAAAGNVTT 1147
QY 1222 KGIVINSQNGONTITGLSNTL-ANVTNDKGSVRTTEGNIKIDEDK--TRAASIVDVL 1278
Db 1148 KEGTINATTGSEVTAQNGTIKGNITSSQNVVTATE--NLVTTENAVINATSGTVNIST 1205
QY 1279 AGFNLOGNGEAVDVSVDYTVNF-ANGNT-TTAKVTYDDTSKTSKVVDVNVDDTTIEVK 1336
Db 1206 KTGDIKGGIE-----STSGNVNITASNTLKVSNITQDVTVTADA-----1246
QY 1337 DKKLGVKTTTLTSTGANKFALSNOATGDALV--KASDVIHAHLNLTLSGDIQ-TAKGASQ 1393
Db 1247 ---GALTTTAGST-----ISATTGNANITTKGIDINGKVESGSGSVTLVATGATL 1293
QY 1394 ANNSAGYVDADGNKVIVYDSTDNKYIQAKNDGTVDKTEKADKLVAAQAQPDGTLAOMNV 1453
Db 1294 A---VGNISGNTVITADS-----GKLTSTV-----GSTINGT-----1323
QY 1454 KSVINKQVNDANKKOGINEDNAPVKLEKAASDNKTKNAAVTVGDLNVAQAQTLTTFAGD 1513

Db 1324 NSVITSSQSGD-----IEGUISGN-TVNVVTASTGD-----ITIGNS 1358
QY 1514 TGTAKKLGETLTIKGGQDTNKLNNIGVAGTGDGFTVKLAKDLTLNLSNVNAGGTKID 1573
Db 1359 AKVEAKNGAATLTAESE-----KLTTOTGSSITSSNGQTTLTAKDSSIAGNINAANVTLN 1413
QY 1574 EKGISFVDANGCOAKANTPVLSSANGLDL-----GGKVISNV-----GKG---TKDTPDAAN 1619
Db 1414 TTGLTTLTGGDSKINATSGTLTINAKDAKLDGAASGDRFTVVNATNAGSGNVTAKTSSVN 1473
QY 1620 VO-OLNEVRNLLGLGNDNA-----DGNQVNIADIKKDPNSGSSSNRTVIKAGTVL---G 1669
Db 1474 ITGDLNII-NGLNIISENGRNTVRLRKEIDVKYIQ-----PGVASVEEVIKRVLEKVK 1528
QY 1670 KGKNDTEKATGVQ-VGVDKDGNANGDLNWWVYKTKDGSKALLATYNAAGQNTVYT 1728
Db 1529 DLSDEERETLAKLGSVAVRFEVPNNA-----ITVNTQNEFTTKPSSQVITISEGKACFSS 1582
QY 1729 NNPAEADIRINEQG 1742
Db 1583 NGARVCTNVADDG 1596

RESULT 13
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-728-470-10

Thu Sep 13 14:19:21 2001

Query Match 4.98; Score 512; DB 2; Length 1529;
Best Local Similarity 21.88; Pred. No. 1.2e-22;
Matches 386; Conservative 229; Mismatches 674; Indels 478; Gaps 91;
QY 143 GDLAEGDASIAIGSDDL-----YLPKRLDLKNEPHK-----LIIGHE 180
DB 72 GQVFLNPGIITIGKDAIINTNGFTASTLDISNENIKARNETLEOTKDALAEIVNHG-- 129
QY 181 ILKKIQTSTGDKIKY--RRTRAQGHASTAVGAMSVAQGHFSNAGTYATAEAAVSLAVGL 238
DB 130 ---LITVGRDGSVNLGGVKKVKEGIVSNGGSISLLAQ-----KIIISD 171
QY 239 AAOATQSSIAVGSNAKANAFATAIGGN-----TVYNLGRGVALGFGSOLLDRNDND 292
DB 172 IINPTIYSIAAPENAEINLGDIFAKGGINVRAATIRNKGK---LSADSVSKDKSGNIV 228
QY 293 ASAYVPLGKTLADQYKATRGQSTDFISGN--SNNNSSIRKLIINVGAGSRDQDAVNV 350
DB 229 LSA-----KEGEA-----EIGGVISAQNOQAQGGKLMITG-----DKVTL 263
QY 351 AQLKVEELANRKITPKGDDNNNSVERGLGNTLTIKGDAQTNALTEANIGVVTGNGL 410
DB 264 KTGAVID-----LSKEGETYLGDEREGEK-----NGIOLAKTKTLEKGTI 307
QY 411 KVKLAKELTGLTSVATNKITVNTNNNAEL--OSGGLTFSPITG---TKTKTVYSIDG 466
DB 308 NVS--GREKGGRAIVMGDIALIDGNIINAQSDIAKTGG--FVETSGHDLSIGDDVI--VDA 362
QY 467 LKFTDNSNATKGT---RITKKKIGFAGTNDGVE-----SKPYLDNEKLKVGNS 516
DB 363 KEWLLDPDDVSIEFTLTSGRNNTGENOGYT--TGDGTEKSPKNSISKPTLTNSTLE---QI 418
QY 517 LNSGSLVNNITGKQIQVANGANGIKFATVANNVANTSAVGTARITEKIGPAGTNDGVD 576
DB 419 LRGSYV--NITANNRIYVNS-----SINLSNGSLTLHTKR-----DGVK 456
QY 577 EAPYLPKELKRVGRVETITDGSINAGNKHITGLTNGIANTDAVTIKOLKDAKPTLNAG 636
DB 457 INGDITSNEN---GNLTAKGWSVDV--HKNITLGTGFLN-----IVAGD 496
QY 637 GISINSNNGD-----LVDSSGNITPTPYNISVTKTKLNSNGTSGNKNFVSNAHDNN 688
DB 497 SVAF--EREGDKARNATDAQITAQGTITVKNKDKQOFRENNVISINGTGKGLKFI---ANQNN 552
QY 689 -----SLVTKADLADYLN-----KVNETAADSALPSEK-----VONGD 720
DB 553 FTHKPDGELNIGSIVTINQTTKDV--KWNASKDSYWNVSSLTNTYQKFTFIKFVDSGS 611
QY 721 NSNNAITVGKDTNGKTFNTLKLKGENGVNITNRTATGTVTFGIDQSNGLTTPKLTGSDT 780
DB 612 NSQDLRSSRRSFAGVHFN--GIGGKTFNIGAN---AKALFKL-KPNAATDPKKEPITF 665
QY 781 NGNRLVIEQVPSADGNSNFKNIKGLSPLPSIA---SPSGRNIALG---NTIEEKDKSNA 834
DB 666 NAN-----ITATDSSSVFMDIHANLTSRAAGINMDSINITGGLDFSTSHNRNSNA 718
QY 835 ASI---DDVLNA---GFNLKN-----NGKDKDFVSTYDTPDFIDGNATTAIVTYDEANQT 883
DB 719 FELKDLTINATGNSFLSKQTKDSFYNEVSKHAINSHNLITLGN-----VTLGGNS 773
QY 884 SKVADVNVDER--TIELTGDKKKQLGVK--TIKLTETSTNGNATFTFSTDDHALVRASD 940
DB 774 SSITGNINITKANVTLQADTSNNTGLKKRPTLTLCNISVEGNLSLTGAN-----AN 825
QY 941 IAGNLTAEEIHTTKGTANTALQ---TFTVKKVDENDKADDTNATITVGKDGTS---KV 994
DB 826 IVGNL-SIAED-STFKGEASDNLITGTFT-----NNGTANINIKQ 864
QY 995 NTLKLG-----KNGLDITKDKGTVTGINTQSLKAGDSTTLNNGNLSIKNTASNEQIQ 1050
DB 865 GYVKLOGDINNKGGLNITNASGTOKTIINGNITNEKGD-----LNKNIKADAELQ 916

QY 1051 VGADGVKFAWNVNVGAGIDGTTTRITRDEIGTGTGNSLDKSKPHLSKDGINAGGKIT 1110
DB 917 IGGN-----ISQKEGNLTISSDKVNITNQITIKAGVEGGRSDSS 955
QY 1111 NIOGSEIAKNSHDAVGGKIYDLKTELENKISSTAKTAONSLEHFSVADEQGNNTVSNP 1170
DB 956 EAENANLTIQTKELKLAG---DLNISGFNKAETITAKNGS---DLTIGNASGNDADAKK- 1007
QY 1171 YSSDYTSKISDVITPAGEN---GITTAKVNGVVRVGIDQTKGLT---TPKLTVCNN----- 1220
DB 1008 -VTFDKVKDSKIIST-DGHNVITLNSSEKTSNGSSNAGNDNSTGLTISAKDVTNNVNTSHK 1065
QY 1221 -----NGKGVINSQNGQNTITGLSNTL-ANVTNDKGSVRTTEQGNIIKDEDK 1267
DB 1066 TINISAAAGNVTTKEGTTINATTSVSEVTAQNGTIRKGNITSONVTVTATE--NLVTTENA 1123
QY 1268 --TRAASIVDVSAGFNLOQNGEAVDFVSVDVNF--ANGNT--TTAKVTYDDOTSKTSKV 1323
DB 1124 VINATSGTVNISTKTGDIKGGIE-----STSGNVNITASNTLTKVSNITQDVTVTADA- 1177
QY 1324 YDVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFALSNOATGDALV--KASDIVAHLNLT 1381
DB 1178 -----GAUTTTAGST-----ISATTNANITTKTGDINGKVES 1211
QY 1382 SGDIQ-TAKGASQANNAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKREVAKDKLVAQ 1440
DB 1212 SGSVTLVATGATLA---VGNISGNTVTTITADS-----GKLTSTV----- 1247
QY 1441 AQTDPGTLAQMNVASVINKEQVNDANKKQGINEDNAFVKLEKAAASDNKTKNAAVTVGDL 1500
DB 1248 GSTINGT-----NSVTSSQSGD-----IEGTISGN-TVNVFASTGD- 1283
QY 1501 NAVAOTPLTTFAGDTGTAKKIGETLITKGGQTDNTKLTDDNNIGVVAGTCDGFTVKLAKDLT 1560
DB 1284 -----LTIGNSAKVEAKNGAATLTAESG-----KLPTQGSSTSSNGGOTTTLTAKDSS 1331
QY 1561 NLSVSNAGTKIDEKGISFVDANGAKANTPVLSANGLDL-----GGKVISNV---G 1609
DB 1332 IAGNINAANVTNTTGTTLTTGDSKINATSGTLTINAKDAKLDGAASGDRVWVATNASG 1391
QY 1610 KG---TKOTDAAVQ-OLNEVRNLLGLGNDNA-----DGNQVNIADIKDPNSGSSNR 1659
DB 1392 SGNVTAKTSSSVNITGDLNTI-NGLNIISENGRNTVRLRGKEIDVKYIQ-----PGVASVE 1446
QY 1660 TVIKACTVL---GGKGNNDTEKLATGGVQ-VGVKDKDNGANGDLSNVWVKTKQKSGKALL 1715
DB 1447 EVTEAKRVLEKVKDLSDEBERETLAKLGVSARFVEPNNA-----ITVNTONEFTTKPSS 1500
QY 1716 ATYNAAGOTNYVTNNPAEAIDRINEOG 1742
DB 1501 QVVTISEGKACFSGNGCARVCTNVADDG 1527

RESULT 14
US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

1660 TVIKACTVL---GGKGNNDTEKATGGVQ-VGVKDGKNGDLSNVVVKTKQSGSKALL 1715
1447 EVIEARVLEKVKDLSDEBRETFLAKGVSAVRFPNNA-----ITVNTQNEFTTRPSS 1500
1716 ATYNAAQGTNYVNNPAEADRINPQG 1742
1501 QVTISGKACFPSSGNGARVCTNVADGG 1527

RESULT 15
US-08-038-682-4
Sequence 4, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GENE III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
City: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-4

Query Match 4.5%; Score 469; DB 1; Length 1477;
Best Local Similarity 21.3%; Pred. No. 4, 4e-20;
Matches 399; Conservative 244; Mismatches 665; Indels 568; Gaps 91:

QY 1 MNHIYKVFENKATGTTFMAVEAYAK--SHSTGGSCATQGVGSVRTLSFARIALAVLVIG 58
DB 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGEKSEKPARM-KVRHLAIKPLSAM-LLSLG 58
QY 59 ATLNGSAYAQIITKTEIGTQNTKINNTLKGDALATGEASIAFGSLSKAQSQAIAIGSVK 118
DB 59 VT-----SIPQSV-----LASGLQMDVVHGTATM-----QVDGNTIIRNSVD 97
QY 119 P-----DPN-----NGSNGVNGSHAKGNESIAIGGOVLAEGDA-----SIALG 156
DB 98 AIIWKQFNIDQNMVQFLQENNSAVFNRTVSNQISQLKGILDSNGQVFLINPNGITIG 157
QY 157 SDDL-----YLPKNLDLKNFEHLIHKIQTSTGDKIKYRRTAQQGHASTAVGAM 211
DB 158 KDAINTNGFTASTLDISN-----NIKARN-----183

212 SYAQGHESNAFCYATAEAAYSIAVGLAAQAATKQSSIAVGSNAKANAFAATAIGNT--- 268
184 -----FTEQTKDKA-----LAEIVNHGLITVKGDSYN-----LIGCKVKNE 221
269 -VVNL-GRGVALGFGSILDRONNTDASAVVPLGKTLADQYKATQCGDSTDFISI-GNSN 325
222 GVISVNGGSISLLAGOKITISDIINPTIY-----STAAPENEAVNLG-----DIFAKGNN 274
326 NNNSSIRRKIIINVAGSRDTDAYNVAQLKVLBEALNRKITFKGDDNNNSV-----ERG 380
275 VRAATIRNQ-----GKLSADSVS-----KDKSGNIVLSAKEGEAEIGGVLSAQNOQA 321
381 LGNTLTIKGDAQTNALTEANIGVVTDCNGLK-----VKLAKELTGLTS 423
322 KGGKLMITGDKVT-----LKTGAVIDLSGKEGGETVLGGDERGEKNGIOLAKK-----YS 372
424 VSATNKITVSWNTNNNAELOSGLTSPITNGTKDTKYISIDGLKFTNDSNSTAKGTTR 483
373 LEKGSTINVSCK-----EKGG--FAIVG-----DIALIDGNIINAQSGSDIAKTG-- 415
484 ITKKKIGFAGT-----NDGVDESKPYLDNEKLKVGNSLTNSGSLTVNTTGNKQIQV 535
416 -----GEVETSGHDLFIKDNAIVDAKEWL-----LDFDNVSINAEADPLFNNTGINDFPT 465
536 G-----ANGIKFATVANNAVNTSATVGTARITTEKIGFAGTND 573
466 GTGEASDPKKNSELKTLTNTTISNYLKNATMTAITSRKLTVNSS-----INI 514
574 GVDEQAPYLDKERLKVGRVEITTDSGINAGN-----HKITGLTNCIANTDAVTI 622
515 GSNHL-ILHSGKQGGGVQIDGITSKGGLTIYSGWVDVHKNIITLDQGFLLAITAASV 573
623 -----KQKDKAKPTLNA-----GDGJISINSNGDL-----VDSGSGNIT- 655
574 AFEQGNKAKDA---ANAKIVAQGVTVTTGEGKDPANNVSLNGTGKGLNISSVNNLTH 630
656 --TPTYNISVTKTLNSNGTSGNNKFSVNAHNNLSVTAKDLADYLNKVNETADALSPLS 713
631 NLSGTINISGNIT---INQTRKNTSYQTSKSHSNVYALNL-----ET--GANFT 677
714 FKVQNGDNNNAITVGKDTNGKTFNTLKLKGENGVNIITNRTATGT-VTFGIDQSGNLTTP 772
678 FIKYISSNSKGLTTOYRSSAGVNFN-----GVNG-NMSENLKEGAKVNFKLPNMMNTS 731
773 K-----LTVGSDTNGNRLVIEQVPSADGNSTKNIIGKSLPTLPSIASPSGRNIALGNTI 826
732 KPLPIRFLANITATGGSVFDDIYANHSG-----RGAELKMSIINISNGANFTLNSHV 784
827 EEKD--KSNAASIDDDVLNAGFNKKNKGKDKDFVSTYDVTDFIDGNATTA----- 873
785 RGDDAFKINKDLTINATNSFSLR---QTKD-----DFYDGYARNAINSTYNSILG 833
874 -TVTYDEANOTSKVAVDVNVDE-KTIELTQDNGCKKQGV--KTIKLTETSTNGNATFTST 929
834 GNVTLGQNSSSSITGNTITIEKAANVTLEANNAPNOQNIIRDVIRKGLSLLVNGSLTGE 893
930 DDDHALVKASDIAGNLNLTAEIHTTKGTANTALQITFTVKVVDENDKADDTNATITVGKDG 989
894 N-----ADIKGNL-TISESA-TFKGKTRDTL-----NITGNTNNGTAENI 933
990 TSGKVNLTCLKGKNGLDKDKTQDVTFFGTINTQSLKAGDSTTLNNGG-LSIKNTASNEQ 1048
934 TQGVVKGNTVNDGDLNI-----TTHAKRNQRSIIGGD--IINKKGSUNITDSNDAE 984
1049 IQVGADGVKFAWVNVGVVAGIDGTTTRITRDEIGTGTNGSLDKSKPHLSKDGINAGKK 1108
985 IOIGGN-----ISOKEGNLT-----ISDKINI-TKQ 1010
1109 ITNIOGSEIAKNSHDVTKGKIYDLKTELENKISSITAKTQNSLHSEFSADEAGNNFTVS 1168
1011 IT-IKKGIDGEDSSSDATSNANLIKTG---ELKLTEDLSISGFNKAEITAKDGRDLTIG 1066


```
QY 1169 NPYSSYDTSKTSVITFAGENGITTKYNGGVVRVIGIDTKGLTTPKLTGVGNNGKGIVIN 1228
Db 1067 N---SNDGNSGAEAKVTTFNNVKDSKISADCHNVTLS-----KVTSSNG-GRESN 1115
QY 1229 SQNGQNTITGLSNFLANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLGAGPNLOG-NG 1287
Db 1116 SDND---TGLTITAKNVEYNK-DITSUKVNTITASEKVT-----TTAGSTINATNG 1162
QY 1288 EAVDFVSTYDTVNFANGNTTAKVTYDDTSKTSKVYVDVNVDDPTIEVKDKKLGKVTTL 1347
Db 1163 KASITTKTGDIISGTISGNTVSVSATVDLTTKSG-----SKIEAKSGEANVTSATG 1212
QY 1348 TSTGTGANKFALSNGATGDALVKASDIVAHLNLTLSGDIQTAKGAS-QANNSAGYVDADGN 1406
Db 1213 TIGGT-----ISGNTVNVNAGDLTVGNGAEINATEGAATLTATGN 1254
QY 1407 KVIYDSTDKYQAKNDGTVDKTKEVAKDKLVQAQTPDGTLAQMNKSVINKBQVNDAN 1466
Db 1255 TL---TTEAGSSITSTKGQVD-----LLAQ----- 1276
QY 1467 KQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNAVAQPLTFAGDTGTAKKLGTLT 1526
Db 1277 -----NGSIAGSINAA--NVTLN---TTGTLTTVA-----GSDIKATSGTLV 1313
QY 1527 IKGCQTDTNKLTNNIGVWAGTDGFTVKLAKDLTNLSNVNAGGT-KIDEKGISFVDANGQ 1585
Db 1314 INAKDAKLN-----GDASG-----DSTEVAVNASGSGSVTAATSSSVNITGD 1356
QY 1586 AKANTPVLSANGLDGGKVISNVGKGT-----KDTD-----AANYQQLNEVRNLLCLG' 1633
Db 1357 LN-----TVNGLN-----IISKDGRNIVRLRGKEIEVKYIQPGVASVEEVIEAKRVLEKV 1406
QY 1634 NDNADGNQVNIADIKDPNSGSSNRNRTVIRAGTVLGGKGNNDEKLGATGGVQGVGDKDGN 1693
Db 1407 KDLSEERETLAKL-----GVSAREVEPNNTITVNTQNEFTTR-PSSQVLIISGRACF 1459
QY 1694 ANGDLNWWVKTKQDG 1709
Db 1460 SSGNGARVCTNVADGG 1475
```

Search completed: September 13, 2001, 12:38:28
Job time: 48224 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:39:59 ; Search time 76.31 Seconds
(without alignments)
2049.356 Million cell updates/sec

Title: US-09-361-619-9
Perfect score: 10356
Sequence: 1 MNHIYKVFKNKATGFMAVA.....NGSADTOGHVGAAGAGHFH 2053

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1191.5	11.5	2059	D82671	surface protein XF
2	971.5	9.4	1588	A86036	probable adhesin z
3	704.5	6.8	1190	A82615	surface protein XF
4	623.5	6.0	5188	B85547	probable RTX famil
5	608	5.9	2249	A41477	190K surface antig
6	591.5	5.7	2240	B71704	cell surface antig
7	553	5.3	2514	F81045	hemagglutinin/hemo
8	542.5	5.2	2660	E85822	probable invasio z
9	540.5	5.2	4919	T31105	hypothetical prote
10	526.5	5.1	2383	D64962	probable membrane
11	526	5.1	3029	S76109	hypothetical prote
12	513	5.0	4152	T31102	filamentous hemagg
13	512.5	4.9	2703	H81193	hemagglutinin/hemo
14	499.5	4.8	2020	C83399	ABC-type transport
15	487	4.7	5627	C83339	hypothetical prote
16	484.5	4.7	2893	A64556	toxin-like outer m
17	464.5	4.5	2154	A83068	hypothetical prote
18	462	4.5	2232	T34434	hypothetical prote
19	461.5	4.5	13288	T03099	mucin, submaxillar
20	460	4.4	1477	B43855	high-molecular-we
21	460	4.4	2468	A83412	hypothetical prote
22	459	4.4	2055	T31110	extracellular matr
23	456.5	4.4	1643	D71630	outer membrane pro
24	455	4.4	1336	A43855	high-molecular-we
25	453.5	4.4	1577	A83140	hemolysin A precu
26	452.5	4.4	1651	JC1340	outer membrane pro
27	449	4.3	2902	C71953	toxin-like outer m
28	446	4.3	3194	D71917	toxin-like outer m
29	439.5	4.2	1902	S06997	lactocepin (EC 3.4

ALIGNMENTS

RESULT 1

D82671

surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: D82671

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82671

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2059 <SIM>

A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1529

Query Match 11.5%; Score 1191.5; DB 2; Length 2059;

Best Local Similarity 23.4%; Pred. No. 6.9e-38;

Matches 548; Conservative 362; Mismatches 771; Indels 657; Gaps 101;

QY 19 VARYAKSHSTGGSCATQGVSVRTLSFARIALAVLVIGATLNGSAYAOITTKIEGQ 78

Db 76 VTGYARFFGSPG--TAAEQOQASRNL-----TLGGSLY----VNSGPGV 114

QY 79 TNKINNTLKGDALATGASIAFGS---LSKAQGSOAIAIGSVKPDNNGSNVGSNAK 135

Db 115 NDVLNKTY-----STMGSVITMNVAGTNAIAIGSAQSSAADALKASLATKASG 164

QY 136 NESTATGGDVLAECDASIAIGSDDLPLPKNLDKNEFKLTHGHEILKIKTQSTPDGKIY 195

Db 165 ARAIAIGAKASADGVDIVALGS-----186

QY 196 RRTRAQGHASTAVGMSYAGQHFSNAFGTATATAAAYSLAVGLAAQATKOSSTAVGSNAK 255

crystalline surfac
cell wall-associat
lactocepin (EC 3.4
probable hemagglut
lactocepin (EC 3.4
probable PpE prote
plasma membrane-as
fat protein - Syne
RTX toxin RtxA VC1
toxin-like outer m
hemagglutinin-like
outer membrane pro
surface-associated
hypothetical prote
cellulose-binding

Db	1090	NGSOLNAVQVQASQPVTFGT-----NEGAVKRSLGOSWVIS-----GESSTAGTY	1134
Qy	1186	AGENGITTKVKNQVVRVIGDQTKLGLTTPKL--TVGNNGKGVINSQNGQNTITGLSNTL	1243
Db	1135	SGGN-LKSVVDEAAGRIHLQLA---DSPKFGVWVNNCGK-----ISCVT---	1175
Qy	1244	ANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLASAGFNLQNGENAVDFVSYDYDVFNFAN	1303
Db	1176	-----AGTEETDAV---NFSOLKSISTAVDQGWLTASGANGSKVASGGTVDLKN	1222
Qy	1304	--GNTTAKVYDDTSKTSKVVDVND-----DITTEVKDKKLGVKVTTLTSTGTG	1353
Db	1223	TDGNLTISK-----SGDSNDVFNLSKDKFVDMTSGTTVWND---GVKVGSDVALGTT	1274
Qy	1354	ANKFALSNOATGDALVAKASDIVAHLNLTLSGDIQPAKASQANNSAGTVADGNKHYDST	1413
Db	1275	GLTIANGPAVTASGIDAGSKVISHV-----AAGAVSETSTDAVNGSOLNAVQVQAS	1325
Qy	1414	DNKYQAKNDGTVDKTEVAKDKLVAQAOTPDGTLAQMNVKSVINKE-----QVNDANK	1467
Db	1326	QPVTFGT-NEGAVKRS--LGOSVVISGESSTAGTYSGNLSKVVDAAAGTTHLQADSPK	1382
Qy	1468	KOGINEDNAFVKGLEKAASDNKTNAAVTVGDLNVAQAOTPLTFAGDTGTTAKKLGETLTI	1527
Db	1383	PGNVINNG---GKISGVTAGTEETDAVNFSQLSKI-----STAVDOGWTLTA	1427
Qy	1528	KG-----GOTDTNKLTDNNIGVAGTGDFTV--KLAKDLTNLSVYNAGGTKIDEKGI	1577
Db	1428	SGANGSKVASGGTVDLKNTDGNLTISKGDSNDVFNLSLSEDLKE-KSTVTVGNTQLDKGV	1486
Qy	1578	S-----FVD-----ANGQAKNTPV-----	1592
Db	1487	KVSNVLLDSNELVITSHSSTSVKTLANGSVNRTVNVGCVNIDVNVVVDLGLSIV	1546
Qy	1593	-----LSANGLDLGGKVISNVGKTDTDAANQOLNEVRNLLGLG-----NMDAGNQV--	1642
Db	1547	GGASLTLSGINAGSHKITVNTAGTEDTDVNFSQLKSVEAVDKGWTLTASGANGSKVVS	1606
Qy	1643	NIADIKK-DPN---SGSSSNRTVI-----KAGTVLGGKGNNDTEKLTATCGVQGVDK	1690
Db	1607	GCTVDLNTDGNLAISKGDSNDVFNLSKDFKVDVETAG-----NTVVNTDGVKVGSDV	1661
Qy	1691	DGNANG-----DLSNVWVKTDGSKKALLATYNAAGTNYVTNNPA---EADIRINEQGI	1743
Db	1662	SIGAMGLFIANGPVSATSGFNAGDK---VISHVAVGMADTDVAVNSQLKQAVQSVTVKAT	1718
Qy	1744	RFEHVNDGNQBPVVQGRNGIDSSASGKHSAVGAIFQAKADGEAAVAIGRQTOA--GNOSIAI	1802
Db	1719	RYSTNDGG---TQGGNYDGDGATGSKAIAAGVGTQASGEGAAVGSAGAAASGKGSTAI	1774
Qy	1803	GDNAQATGQOSIAIGTG-----NVVACKHSG-----AIGDPSTVKADNSYSVG	1845
Db	1775	GRNATASAGSVALGDGAKDGGRAESYTGKSGVQNNVTGTVSVGDAKAGETRSINVA	1834
Qy	1846	NNNOFTDAT-----QT-----DVEGV--GNNTVVTES---NSVAL	1875
Db	1835	DAKEAMDAVNLQDLDAVAKSNLQDDMRHEINNTIEDVFKITGDSASSVKGCMGVNMAI	1894
Qy	1876	GNSAISAGTHAGTQAKSDGTAGTTTTAGATGTGVKFGAGOTAVGAVSVGASGAERRIQN	1935
Db	1895	GTNAAYS-GTESVALGKNTVNSADNAVAIG--NGSVADRA-----NSVSVGSGGSEROVTN	1947
Qy	1936	VAAGEVSATSTDAVNGSOLYKATQSTANATNELDHRHONENKANAGISSAMAMASMPQA	1995
Db	1948	VAAG---TADTDVAVNSQL---NOGLITAKQYTDGMVGNLRRRTSGGVAALATANLQQA	2001
Qy	1996	YIPGRSMVTGTHNGQGVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGVAGFHF	2053
Db	2002	YVQCRGMTSVGVSSYQOQSAIAVGSVSESJHWFKPSGSANTRSHVGVGAVGVQYQW	2059

Db	187	-----GATA-----GT-----GASSIAIGLNASAV-NGAVAVGGGAL	217
Qy	256	ANA--FAATAAGNTVNLGRGVALFGQSILDRDNTDASAVPLGKTLADQYKATROGD	314
Db	218	VTPDGAVALGNSVASTGKGLS---GYDPKTKTSTDASA-----AWKSTLAA--	263
Qy	315	STDIFSIGNNNSSIRRIKINVAGSRDQDVAVNAQLKVELANRKITFKGDDNNS	374
Db	264	-----VSIQDSSTNLK-TROLSGLAAGTSNTDAVNAQLKVDVDEIASR-----	306
Qy	375	NSVERGLNTLIKDAQTNALTEANIGVYVDGNGLKVKLAKELGTLSVSATNKITVSN	434
Db	307	-----GWNLTASG-----ANSGNAPGSSVDLKNKNTDKNLTITKAIGSNDVQF--	348
Qy	435	TNNNAELQSGGLTFSPITGKTOKTVVYSIDGLKFTNDSNSIATKGTTRITKKKIFAGT	494
Db	349	--NLNKDKV-----VTTLAVGDAALLNTDGIAGLTDV-SLSTTGLA-ITD---GPAVT	393
Qy	495	NGVDDESKPYLDNEKLKVGNSFLNSGLTVNNTGNKQIQGVANGIKFATVANNVANTS	554
Db	394	ASGIDAGSKVISH-----VAAGAVSETSDAVNGSOLNAVQVQAS--QPVTFGTNEGAVKR	447
Qy	555	TVGTARITEKIGFAGTNDG-----VDEQAPYLDKERLKVGRVEIITDGINAGN---	604
Db	448	SLGQSVVISGESSTAGTYSGGNLSKVYDEAA-----GRIHLQADSPKFGNVVIN	497
Qy	605	--HKITGLINGIANTDAVITIKOLDAKAPTLNAGDGISINSNNG-----DLVDSGN	653
Db	498	NGKISGVTAGTETDAVNFSQLKSISTAVDQGWTLTASGNSKVASGGTVDLKNTDGN	557
Qy	654	ITPTYINISVKTKLNSGTSFGNKKFSVS-----NAHDNLSVTAKLADLYLN	701
Db	558	LT-----ISKGDSNDVFNLSKDFKVDGMTSGTTVWVNDGVKVGSDVALGTT	605
Qy	702	KVNETASALPSFKVQNGDN--SNNATVVGKDT-----NGKTFNTLKLKGENVITNR	754
Db	606	GLTITDGPVATSGIDAGSKVISHVAVGVSETSDAVNGSOLNAVQVQASQPVTFGTGNE	665
Qy	755	ATGVTFGIDQSNGLTPKLTGVSDTNGN-RLVIEQ-----VPSADGN	796
Db	666	--GAVKRSLGQSVVISGESSTAGTYSGGNLSKVYDEAAAGRIHLQADSPKFGNVINNG	723
Qy	797	STKNIKGLST-----LPSIAS-----PSGRNIALGNTTEEDKKSNAASI	837
Db	724	KISGVTAGTEETDAVNFSQLKSISTAVDOGWTLTASGNSKVASGGTVDLKNTDGNLTI	783
Qy	838	D-----DVLNAGFNLKN-----GKDKDFVSTYDVTDFIDGNATTAT	874
Db	784	SKSGDSNDVFNLSKDFKVDGMTSGTTVWVNDGVKVGSDVALGTTGLTI--TDGPVATAS	841
Qy	875	-----VTYDEANOTSKVAYD-VNVDK-----KTIELFGDNG--KKQLGVKVTIK	914
Db	842	GIDAGSKVISHVAAGVVSETSDAVNGSOLNAVQVQASQPVTFGTNEGAVKRSLGQSVVI	901
Qy	915	LTETSTNGNATTFSTDDHALVKASDIAGNLTLAE-----IH-----	953
Db	902	SGESSTAG---TVS-----GGLNLSKVYDEAAGRIHLQADSPKFGNVINNG	945
Qy	954	-----TTKGNTANTALOTFTVKVVDENDKADDTNATVVGKDGTSKGVNTLKLKNGLDIK	1008
Db	946	GKISGVTAGTEETDAVNFSQLK-----SISTAVDQGWTLTASGANGSKVASGGTVDLK	998
Qy	1009	TDKDGTVTFGINTQSLKAGDSTTLNNGLSIKNTASNEQIOVGADGVKFAWNVNVVGA	1068
Db	999	NTDGNLTIS-----KSGDSNDV-----VFNLSKDKFVDMTSGT--TVWVNDGVKV	1042
Qy	1069	GID--GTTTRITREIGFTGNSLDKSPHLSKDGIDAGGKKTITNOSGETAKNSHAY	1125
Db	1043	GSDVALGTT-----GLTIANG-----PAVTASGIDAGSKVISHVAAVAGVSETSDAV	1089
Qy	1126	TGGKIYDLKLENIKISSTAKTQNSLHIEFSVADEQGNFTVSNPYSSSYDTSKTSDVITF	1185

probable adhesin Z5029 [imported] - *Escherichia coli* (strain O157:H7)
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A86036
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1588 <STO>
A:Cross-references: GB:AE005174; NID:gl2518349; PIDN:AAG58749.1; GSPDB:GN00145; UWGP:Z50
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z5029

	Query Match	9.4%; Score 971.5; DB 2; Length 1588;
	Best Local Similarity	22.4%; Pred. No. 1.2e-29;
	Matches	Conservative 281; Mismatches 651; Indels 787; Gaps
Qy	1 MNHIYKVFNKATCTPMVAEYAKSHSTGGSCATCGVCSVRTLSFARIAALAVLIVIGAT	60
Db	1 MNKIFKIWNPATCNVTYTSETAKSRG-----KGRSKLLISALVAGMLSSFGAL	52
Qy	61 LN-GSAYAQOITTK-----IEIGOTWNKNINTL--KGDALATGEASITAFGLSLKAOGS	109
Db	53 ANAGNDNGCQVDYGSGSAGDCWVAIOGKAKANTFMNTPSGSSTAVGYDAIEAEOYSALGS	112
Qy	110 QATAIGSVKPPDPNGSNNGNVGSHAKNESIATIGDVLAEADSIATIGSDDIYLPKNLDLK	169
Db	113 K-----THAIGGASMAGFSVASEGDSRSIALGASSYSLGOYSMAL	152
Qy	170 NEPHKLHGHLEILLKKIQSTDGKIKYRPRTAQGHAHVAGMSYAQGH-----FSNAF	222
Db	153 GRYSKAL-----GKLST-----AMGSDSKAEGANAIALGNATKATEIMSTIAL	194
Qy	223 GTATATAEAYSILAVGLAAOATKOSSIATVGSNKAKANAFATATGGNTVVNLGRGVALGFES	282
Db	195 GDATANASKAYSMALGACSSVASENAITAIGAETEA-AENATATCNKAKAKGTNSMANGFGS	253
Qy	283 QILDRODNNTDASAVPLG---KTLADQYKATRG---DSTDIFSIGN-----	323
Db	254 -LADKVNT-----IALNGSQALADNAITAGGKNKADGVDAIALGNGSQSRGLNTIALG	306
Qy	324 -----SNNN-----NSIRRKIIHWGAG---S	342
Db	307 TASNATGDKSLALGNSSSANGINSVALGADSTADLDNTVSGNSSILKRKIVNVKNGAITS	366
Qy	343 RDTDAVNVAOLKLVHEELANKRIKFKGDNNSNVSVERGLNTLTTKGDAOTNALTEANIG	402
Db	367 DSYDAINGSQLYAI-----SDSVAKRLGGGAADVDDGCTVTATPYNL-	408
Qy	403 VVTDGNGLYKLAKELTGLTSVATNKITVSNPT---NNNNAELQSGLTFPSITGKTDK	459
Db	409 ---KNGSK-----NNVGAALAVLDENTLOWDOTKGYSAAHGTSPTSASVIID-	453
Qy	460 TVYSISDLKETDNSNIATGKTTRITPKKIGPAGTWNGDVDESKPYLDNEPKLVKNGSTLNS	519
Db	454 -----VADGTTISAS-----SKDAVNGSQLKATNDNDVEA	481
Qy	520 GSLTVNNTTNKQIQVANGCIKEATVANNVANTSATVGTARI-----TEEKIGFACTNDG	574
Db	482 NTANIAFNISN----TATNTANIATNTNITNLTDVSGDQLADALLWNTEKKAFSAAH-G	536
Qy	575 VDEQAPYLDRERLKVGREITTDOSGINAGNHKITGLTGINTANTDAVTIKOLKDAKPTLNA	634
Db	537 QDTTSKITN-----VRDADLTADS-----TDVNGSQLK-----TT	567
Qy	635 GDGISINSNGDLVDSSGNTTTPTYNISVKTTKLNSNGTSG-----NNKFSVSNADHN	688

Dd	568	NDAVATNTN---IANTNSIA	TNTNI	SNLFTVN	LGEDALKWDKONGVFTA	AHGTETT	625
Qy	689	SLVTAKDLADYLNKVNETA	SALPSFKYQNGD	NSNAITVGKTNGTNTKLKG	ENG V	748	
Dd	626	SKIT----	NVKDGD-----	LTTG-STDAVNGSQLTND	AVA	657	
Qy	749	NITTNRATGVTFEGIDOS	NGLTTPKLT	VGSOT-----	NGNLVIEQVP	SADGNSTKNII	802
Dd	658	TNTNTIATNTNI-----	SNLFTVN	LGEDALKWDKONG-----	VFAHGHNTASKI	706	
Qy	803	KGLSPTLP	SIASPSGRNALGNT	IEEKDSNAASIDDVLNAG---	FNLKNGKDKDFVST	860	
Dd	707	TNI-----	LDGTV-----	TATSSDRAINQSOLDJSSN-----	IATY	737	
Qy	861	DTVD	IDGNATTATVYEANOTS	KVAYDVNVDEKTIELTGDNKKOLGV---	KTIKLTE	917	
Dd	738	----	FGNAS-----	VNTD-----	GVTGPTYKIGE	759	
Qy	918	TS--	TNGNA-----	TFSTPDDDHALVKASDI	IAGNLNTLAEBIH	TKGTANTALQTFTVKK	970
Dd	760	TNYNVGDALA	AAINSFSFSLGDALLMDA-	TAGKFSAK-----	HGTNGDASVI-----	806	
Qy	971	VDENDRADOTNAITVCKD	GCTSGKVNTLK	LGKN-----	GLDIKTDKDGTV---	TGIN	1020
Dd	807	---	TVDADGEIS-----	DSSDAVNGSQLHGVSYYVDALGGGAEVNADG	ITAPTATIA	858	
Qy	1021	TQSLKAGADSTLLNN	GLSIKNTASNEQIOVGADCV	KPFAMVNVGVGAGIDGTT	RTRDE	1080	
Dd	859	NADYD	NVGDA-----	LNAITDDLALMDAD-----	AGENG-----	890	
Qy	1081	IGFTGTNGSLDKSPHL	SKDGINAGCKKITNTQSGE	IAKNSHDVATGKIYDLK	TELENK	1140	
Dd	891	-AFSAAHGK-DKT-----	ASVITNVANGAISAA	SSDAINGSQLTYTNKYIADA	936		
Qy	1141	ISSTAKTAQN---SLHEFS	VADBOGNFTVSNPYSSYDT	SKTSDVITFAGENGITTKVNK	1197		
Dd	937	LGGDAEVNADGTTIAPT	YTIANAEYN-----	VGDALDALDALLM-----	978		
Qy	1198	GVVRVGIDOTKGLTTP	KLVTVGNNGKGI	VINSONGONTIGLSNTLANVT	NKDGSVRTTE	1257	
Dd	979	-----	DETA-----	NGGAGAYNASHDGKASI-----	ITNVAN-GSI-----	1008	
Qy	1258	OGNI	I IKEDKTRAASIVDWLS	AGFINLQGNGEA	AVFVSYDTVNFANGNTT	TAKVYDYDTS	1317
Dd	1009	-----	SEST-----	DAVNGSQLNATNMIL-----	E	1029	
Qy	1318	KTSKVYDV-V	NVDDTTIE-----	VKDKLGVKTTTLSTGTGANK	FALSQAOTGDA	1367	
Dd	1030	QNTQIINQLAGNTDATY	IOENGAGINYRTDGD	LAFANDASAQGVGATGYNS	VAKGS	1089	
Qy	1368	LV-----	KASDIVAHLNLTLSGDIOT---	AKGASOANNAGVYVDAGNK	VIYDSTDNKYYQ	1419	
Dd	1090	SVALGQGSYSDVDTG	TALGSSSVSRVIAKGRDTS-----	ITENGVVIGYD	TTDGBELLG	1144	
Qy	1420	AKNDGT	VYDKTKEVAKDKLVAQAQ	PTDGTTLAOMNVKSVINKEQVNDANK	KQGINEDNAFVK	1479	
Dd	1145	ALSIG-----	-----	DDGKYRQIIN-----	1159		
Qy	1480	GLEKAASDNKTKNA	AVTVGDL-----	NAVAAQTPPLTF---AGDTGTTAKLGET	TLTKGQOTD	1533	
Dd	1160	-----	VADGSEADHAVTRQLQ	AIACAVATPTKYFHANSTEEDSLAV	GTDSLAMGAKTI	1214	
Qy	1534	TNKLTDNNIGV	AVAGT-----	DGFVTKLAKDLTLNLSNVNAGG	TKIDEKG---ISFVDANG	1584	
Dd	1215	VN--	GDKGIGIGYGAVVDANALING	IAIGSNAQVIHVNSTAIGNSGSTTTR	GRAQTNITYAYNM	1272	
Qy	1585	QAKANTPVL	SANGLIDLGGKVI	SNVGKGTKXDTDAANVOQLNEVRNL	LGLNDNADGNVNI	1644	
Dd	1273	DAPONSVGEFSVGSAD	GORIITNVAAGSADTDAVNV	QOL-----	-----KV	1313	
Qy	1645	ADTKDPNSGSSNR	RVIRKAGTVLGVGKGNNDTE	KIATGVQVGVDKGNANGDL	SNWVWK	1704	
Dd	1314	TDAQVSONTQSI	JTN-----	LDNRVFNLDLSRVT---NIENGJ-----	GDIV-----	1350	

Thu Sep 13 14:19:27 2001

1705 TQDGSKKALLATYNAAAGQTYNNPAAEADIRNEQGIREFHVDNQEPVVGRRNGID 1764
 1351 -----TTGSKTKYFKTN-----TDGVD 1366
 1765 SSAGSKHSVAIGFQAKADGEAANAIGQTOAGNOSTAIGDNAQATGDOQIAIGTGNVAG 1824
 1367 ASAGKDSVAI-----GSGSIAADN-----SVALGTGSV--- 1396
 1825 KHSAGIDPSTVKADNSYSVGNNNQFTDQTQDFVGVGNNTVTSNSVALGSNSAISAG 1884
 1397 -----ATEENTISVGSSTNORRITNVAAGKNA----- 1423
 1885 THAGTQAK-KSDGTAGTTTAGATGTVKGFAGQTAAGVAGSVGASGAERRIONVAAGEVSA 1943
 1424 TDAVNAQLKSEAGGVRYDTKADGSD--YSNITLG---GNGGTTTRISNSVAG--V 1474
 1944 TSTDVANGSOLYKATQSTANATN---ELDHRIHQENKANAGISSAMAMASMPQAYIPG 1999
 1475 NNDVNVNAQLKQSVQETKQYTDQRMVEMDNKLSKTESKLSGGTASAMAMTGLPQAYTPG 1534
 2000 RSMVTGGIATHNGCAVAVGLSKDSNGOWEKGADTQGHVGAAGAGFHF 2053
 1535 ASMASIGGGTYNGESAVALGVSWANGRWYKLOGSTNSOGESYAALGAGIQW 1588

1705 TQDGSKKALLATYNAAAGQTYNNPAAEADIRNEQGIREFHVDNQEPVVGRRNGID 1764
 1351 -----TTGSKTKYFKTN-----TDGVD 1366
 1765 SSAGSKHSVAIGFQAKADGEAANAIGQTOAGNOSTAIGDNAQATGDOQIAIGTGNVAG 1824
 1367 ASAGKDSVAI-----GSGSIAADN-----SVALGTGSV--- 1396
 1825 KHSAGIDPSTVKADNSYSVGNNNQFTDQTQDFVGVGNNTVTSNSVALGSNSAISAG 1884
 1397 -----ATEENTISVGSSTNORRITNVAAGKNA----- 1423
 1885 THAGTQAK-KSDGTAGTTTAGATGTVKGFAGQTAAGVAGSVGASGAERRIONVAAGEVSA 1943
 1424 TDAVNAQLKSEAGGVRYDTKADGSD--YSNITLG---GNGGTTTRISNSVAG--V 1474
 1944 TSTDVANGSOLYKATQSTANATN---ELDHRIHQENKANAGISSAMAMASMPQAYIPG 1999
 1475 NNDVNVNAQLKQSVQETKQYTDQRMVEMDNKLSKTESKLSGGTASAMAMTGLPQAYTPG 1534
 2000 RSMVTGGIATHNGCAVAVGLSKDSNGOWEKGADTQGHVGAAGAGFHF 2053
 1535 ASMASIGGGTYNGESAVALGVSWANGRWYKLOGSTNSOGESYAALGAGIQW 1588

RESULT 3
 A:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: A82615
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82615
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1190 <STM>
 A:Cross-references: GB:AE004017; GB:AE003849; NID:99107083; PIDN:AAF84783.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Franco, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1981

Query Match 6.8%; Score 704.5; DB 2; Length 1190;
 Best Local Similarity 22.9%; Pred. No. 1.2e-19;
 Matches 360; Conservative 179; Mismatches 508; Indels 525; Gaps 65;
 QY 624 QLKDAKPTLNAGDGSINSNNGDLVDSSGNITPTTNYI-----SVKTKLNSNGT 673
 DB 2 QIHTALPMVRLGG-----IEGVDPFLPKYKIGQSLQHAWMTSAASSKKGTOPR 50
 QY 674 SGNNKFSVSN-----HDNNSL--VTAKDLA---DYLKNVETADSAALPSFKVONGDNSNA 725
 DB 51 RSNAMTAKRSKALHRRQRHLVLLVLAASTGYTKGAAQV-----YVNSDSTENC 102


```

      868  -----GATGTRSIAGVGTLASAEGATAGSG 894
      1793  TQA-GNQSTAGNAOATGQSTAIATG-----NVVAGKHS-----AIGDPS 1834
      895  AAASRGSTAGRNAVASAGSVALGDGARGAESYTGKYSGLQNNVTGVSVDAS 954
      1835  TVKADNSYSGVNNNOFTDAT---QTDVFG-----VGNNI-----TVTESNSVALGSN 1878
      955  KGETRTVSNDAKEATDANLRLQDRVAQDANRYVDNKIESLSEBQTFVKVNSL---NN 1011
      1879  SA--ISAGTHA-----GTAQRKSDGTA-GTTTTAGATGTVKFGAGTAVG-----A 1921
      1012  SATPIAAGVDATAIGVATGASGADSIAMGNKASADNAV-----AIGNHSVADRANT 1064
      1922  VSVGASGABERRTONVAAGEVSATSDAVNGSOLYKATQSIANATNDELRIHONENKANA 1981
      1065  VSYGSAGSERQVTNVAAG---TADTDAVNSQL---NOGLITAKQYTDGVGVSLRRDITDG 1118
      1982  GISSAMAMASMPQAYTPGRSMVTGGTATHNGOGAVAGLSKLSDNGOWFEKINGSADTQG 2041
      1119  GVAALATLANLPQAYTPGRGMTSVGSSYRGQSAIAVGVSSVSESGRWVFKFSGSANTRS 1178
      2042  HVGAAYGAGGFH 2053
      1179  QVIGAGVGVQW 1190

RESULT 4
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoculis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20615

Query Match 6.0%; Score 623.5; DB 2; Length 5188;
Best Local Similarity 22.6%; Pred. No. 1e-15;
Matches 526; Conservative 287; Mismatches 846; Indels 673; Gaps 126;

QY 32 SCATQGVSVRTLSFARIAALAVLVIGATLNGSAYAAQIITTKIEIGQTNKNNITLKGDD-- 89
DB 1497 SVPTGLASLTSASYTNAS---VSDKARNASASATHNLT--VDLAAPVVTINTVAGDDI 1550
QY 90 --ALATGEASIAFGSLUSKAGGQAIAGSVKPPNNGSNCNGVS-----HAKGNESIA 140
DB 1551 INATEHGQAIIISGATGATTGNTSV-----TIGTTTYYTVLDANGNWSIG 1597
QY 141 IGGDV---LAEGDAST-----AIGSDDLVLPKNLDDLKNEFKLIH 177
DB 1598 VPASVISAQAQGVDTITATVTSAGNSGTASHTVTVVALGAPVLAI-----NTA 1646
QY 178 GHEILKIQITSDGKIKYRRTRAQG--HASTAVGAMSY-----AQHFS---NAFGTYATA 228
DB 1647 VDDIINAEKGADLAITGTSNQAGTQIIVTLNGQNYTTTADAGNWSVTVPASRVYALG 1706
QY 229 EAYSIAVLGAQAATKQSITAVGSNAKANAFAATATGGNT-----VYNLGR-QVALG 279
DB 1707 EATYT--VTAATAADADGNSGASHNQVN-----TALPGVTINVVATDDIINAAEAGEQT 1760
```

```

QY 280 FGSQILDRONNTDASAVVPLGKTLAD-QYKATRGD---STDI-----FSIGNSN-NNNS 329
DB 1761 ISQGV-----TGAAGADGTVTTLGGATYATATQANLSWSVDVPASALQELNGELTISA 1814
QY 330 SIRRKIINVAGSRD--TDANVNAQLKVLBELANRKITFKFGDGNNSNSVBERG----- 380
DB 1815 SVTNSVGTNGTGREITIDANLPGLR-VDIVA-----GDDVNIIEHQALVITGS 1864
QY 381 -----LGNLTITIKGDAQT-----NALTENIGVVTDCNGL 410
DB 1865 SSGLAAGSNVTLTINGQTYVAAVLADGTSVGPVAVDVSAMPAGSVTIAASGSTSAGNPV 1924
QY 411 KV--KLAKELTGL-TSVSATNKITVSTNNNAELQSGGLTSPITGKTGKTDKTVYSIDGL 467
DB 1925 SVTHPVTVDLSAVAVSINAITADDVINAEEKGAALTLSGST-----SCVEAGQTVTVTFGG 1980
QY 468 KFTNDSNSTATKGTTRITKKKIGFAGTNDGVDSKPKYLDNEKLVGNST-----LN 518
DB 1981 K--TYSATVAANGSWSTSVPADMAALRDGASQAQSVSNV--GNSATTHAYSVDAS 2035
QY 519 SGLTVNNTTGNKQIOVGANGIKPATVANNVANTSATVGTARITEEKIGFAGT--NDGV- 575
DB 2036 APTVTINTIAGDDILNAAEAGAALTITGSSTAEGQTV--TVTLNGTNTGTVTQDGSW 2092
QY 576 DEQAPYLDKERLKVGRVEITTDSCINAG-----NHKITGLT-----NGIANTDAYTIK 623
DB 2093 SVSVPASADLSTLTASXYTVNAAVSDRAGNPASVNNHNTVDTSVPVVTVINTVAGDDVFNAT 2152
QY 624 QLKDAKPTLNAGDG-----ISINNN-GDLVDSGNTTPTYNISVKTTLN--SNG 672
DB 2153 EHAQAQIISGSATGATGTSVTVTIGTNTTFTVLDASG-----WSGVGPASVVSALANG 2207
QY 673 TSGNNKFSVSNADHNNLSLVTAKDLADLYLNKVNETAASLP--SFKVQNGDSNNA----- 725
DB 2208 TVTINA-SVTDAGNSGSAHQ-----VTVNTGLPTITFNALSGDNLNADSKGQ 2256
QY 726 -ITVGDKTNGKTEN--TLKLGKENGWNITNRTATGVTGFIQDSNGLTTPKLTVGSDTN 781
DB 2257 PLTISGSGTGLATGAQVTVTLNGHN-YSATDT-ASGNWT-----LTPVPSDLAALQG 2306
QY 782 GNRLVIEQVPSADGN---STKNIIKGLSPTLPSIASPSGRNIALGNTIEEKDSNAASID 838
DB 2307 ANYTVSASATSAAGNTASSQANLL--VDSGLPDV-----TI-----NTVAGD 2346
QY 839 DVLNAGFNLNKNGKDKDF-----VSTYDTPVDF-IDGNATTATVYDEANOTSKVAYDVN 891
DB 2347 DLIINAA---EAGADQTI SGVWTRAAAGDTVTVTGLGNTYTATV-----QSNLSMSVS 2395
QY 892 VDEKTIETLGDNGKKQGVKTKIKITETSTNGNATTFSTDDHALVKASDIAGNLTLAE 951
DB 2396 VPTADLQALGN-----GDLTITASVTNANGTSGSPRD-----ITIDANLPGL--- 2438
QY 952 IHTTKGTANTALQTFYVKKYVDENDKADDTNATITVGKDTSGKVNTLKLKGNGLDIKTDK 1011
DB 2439 -----RVDTVAGDDIVNSIEHQ-----ALVITGGS----- 2464
QY 1012 DGTVTFTGINTQSGIKAGDSTTLNNGLSIKNTASNEOIQVAGDVKFAMVNNGVYAGID 1071
DB 2465 -----SGLNAGAVLTVTINSVAYSAT-----VQADGSMWSVGIPAAVNSAWPA 2506
QY 1072 GTTRITRDEIGFTGTNGSLDKSKPH-----LSKDGINAGGK-----KI 1109
DB 2507 GPLTVEVD--GQSSANNPVSHPFTVDLTPAVAISINTVASDDVINAAEKGNTJLTLSGST 2564
QY 1110 TNIQSGEIAKNSHDAVT-GGKIY-----DLKT-----ELENKISSTA 1145
DB 2565 SGIESQTV-----TVTFGKTYTASVAANGSWSVNVPAADLATLPEGANVQASVSAS 2619
QY 1146 KTAQNSLHEFSVADEQGNFTVSNPYSSYD---TSKTSDDVITFAG-----ENGITTKVN-K 1197
DB 2620 GNSASATHAYSV-DASAPTLTI-NTIASDDILNAAEAGSPLTISGTSTAETGTGTVTTLN 2677
QY 1198 GVVRVGIDQTKG---LTPPKLTVGNNGKGINV---SQNGQNTITGLSLANLVNDKG 1251
```


A:Reference number: A41477; MUID:90354033	
A:Accession: A41477	
A:Molecule type: DNA	
A:Residues: 1-2249 <AND>	
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466	
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi	
C:Keywords: surface antigen; tandem repeat	
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>	
Query Match 5.9%; Score 608; DB 2; Length 2249;	
Best Local Similarity 22.5%; Pred. No. 1.3e-15;	
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;	
Qy	82 INNTLKDALATGEASIFAGSLSKAQGSOAIAIGSVKPPDPPNNGSNVGSN-----132
Db	13 IQOGLKAALFTTSTAIMLSS-----SGALGVATGVIATNNNAFNNYGNWNNWEITAAG 68
Qy	133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176
Db	69 VANGTPAGGPQNMAFTYGGDYTVTTADAADRIKAINVAGT-----TPVCLNI-----116
Qy	177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224
Db	117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFDPADNVTGLGN 168
Qy	225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272
Db	169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223
Qy	273 GRGVALGFGSOILLDRNN-TDASAYVPL-----GKTLADQYKATROGSDTDFSI-----321
Db	224 GAGTATLGGAVIKATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283
Qy	322 -GNSNNNSSTIRKIIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374
Db	284 TGDIGNTS---LATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVVVTCADNTG 340
Qy	375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCGNGLKVLKAKELTGLTSVSAT 427
Db	341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLLQVGVVVKANTINLTNDSAVTET 399
Qy	428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG----480
Db	400 NPVVVTCADNTGNANNGIVTFGNTSVTVG-----DIGNTNALATVNVGAGTATLGGAVI 454
Qy	481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTTGNKQIQV-G 536
Db	455 KATTKTLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494
Qy	537 ANGKEATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578
Db	495 LNG-ALSOQVTGIGNTSLATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVV---550
Qy	579 APYLDKERLKVGRVEIT--TDSGI--NAGNHKIT---GLTNGIAN-----616
Db	551 -----VTGADNTGNANNGIVTFGNTSVTVGDTGNTSLATISVGAGTATLGGAVI 601
Qy	617 -----TDAVTIKOLKDAKPTL-----NAGDGISINSNNGDLVDSGGNI--TTPYN 660
Db	602 KATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLAT 661
Qy	661 ISVKTTKLNSGTSGNNKFSVSNADHNNSLVTAKDLADYLNKVNETAADSALPSEKVGND 720
Db	662 ISVGA---GTATLGGAVIKATTKITNAVSARK-----FTNPVVVTG-----AIDSTG 706
Qy	721 NSNNAI---TVCKDNTNGKTFNTLKLKGENGVNITTNRATGTVTTFGIDQSNQ-----LTPPK 773
Db	707 NANGIVTFGNTSVTVGDTGIGNTAL-----ATNVGAGTATLG-----GAVIKATTK 754
Qy	774 LTVGSD-----TGNRNLVIEQVPSADGNSTKNIK-----GLSPTLPISIASPSG 817
Db	755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLATISVGAG 814

A:Reference number: A41477; MUID:90354033	
A:Accession: A41477	
A:Molecule type: DNA	
A:Residues: 1-2249 <AND>	
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466	
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi	
C:Keywords: surface antigen; tandem repeat	
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>	
Query Match 5.9%; Score 608; DB 2; Length 2249;	
Best Local Similarity 22.5%; Pred. No. 1.3e-15;	
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;	
Qy	82 INNTLKDALATGEASIFAGSLSKAQGSOAIAIGSVKPPDPPNNGSNVGSN-----132
Db	13 IQOGLKAALFTTSTAIMLSS-----SGALGVATGVIATNNNAFNNYGNWNNWEITAAG 68
Qy	133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176
Db	69 VANGTPAGGPQNMAFTYGGDYTVTTADAADRIKAINVAGT-----TPVCLNI-----116
Qy	177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224
Db	117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFDPADNVTGLGN 168
Qy	225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272
Db	169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223
Qy	273 GRGVALGFGSOILLDRNN-TDASAYVPL-----GKTLADQYKATROGSDTDFSI-----321
Db	224 GAGTATLGGAVIKATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283
Qy	322 -GNSNNNSSTIRKIIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374
Db	284 TGDIGNTS---LATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVVVTCADNTG 340
Qy	375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCGNGLKVLKAKELTGLTSVSAT 427
Db	341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLLQVGVVVKANTINLTNDSAVTET 399
Qy	428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG----480
Db	400 NPVVVTCADNTGNANNGIVTFGNTSVTVG-----DIGNTNALATVNVGAGTATLGGAVI 454
Qy	481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTTGNKQIQV-G 536
Db	455 KATTKTLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494
Qy	537 ANGKEATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578
Db	495 LNG-ALSOQVTGIGNTSLATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVV---550
Qy	579 APYLDKERLKVGRVEIT--TDSGI--NAGNHKIT---GLTNGIAN-----616
Db	551 -----VTGADNTGNANNGIVTFGNTSVTVGDTGNTSLATISVGAGTATLGGAVI 601
Qy	617 -----TDAVTIKOLKDAKPTL-----NAGDGISINSNNGDLVDSGGNI--TTPYN 660
Db	602 KATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLAT 661
Qy	661 ISVKTTKLNSGTSGNNKFSVSNADHNNSLVTAKDLADYLNKVNETAADSALPSEKVGND 720
Db	662 ISVGA---GTATLGGAVIKATTKITNAVSARK-----FTNPVVVTG-----AIDSTG 706
Qy	721 NSNNAI---TVCKDNTNGKTFNTLKLKGENGVNITTNRATGTVTTFGIDQSNQ-----LTPPK 773
Db	707 NANGIVTFGNTSVTVGDTGIGNTAL-----ATNVGAGTATLG-----GAVIKATTK 754
Qy	774 LTVGSD-----TGNRNLVIEQVPSADGNSTKNIK-----GLSPTLPISIASPSG 817
Db	755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLATISVGAG 814

A:Reference number: A41477; MUID:90354033	
A:Accession: A41477	
A:Molecule type: DNA	
A:Residues: 1-2249 <AND>	
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466	
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi	
C:Keywords: surface antigen; tandem repeat	
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>	
Query Match 5.9%; Score 608; DB 2; Length 2249;	
Best Local Similarity 22.5%; Pred. No. 1.3e-15;	
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;	
Qy	82 INNTLKDALATGEASIFAGSLSKAQGSOAIAIGSVKPPDPPNNGSNVGSN-----132
Db	13 IQOGLKAALFTTSTAIMLSS-----SGALGVATGVIATNNNAFNNYGNWNNWEITAAG 68
Qy	133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176
Db	69 VANGTPAGGPQNMAFTYGGDYTVTTADAADRIKAINVAGT-----TPVCLNI-----116
Qy	177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224
Db	117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFDPADNVTGLGN 168
Qy	225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272
Db	169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223
Qy	273 GRGVALGFGSOILLDRNN-TDASAYVPL-----GKTLADQYKATROGSDTDFSI-----321
Db	224 GAGTATLGGAVIKATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283
Qy	322 -GNSNNNSSTIRKIIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374
Db	284 TGDIGNTS---LATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVVVTCADNTG 340
Qy	375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCGNGLKVLKAKELTGLTSVSAT 427
Db	341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLLQVGVVVKANTINLTNDSAVTET 399
Qy	428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG----480
Db	400 NPVVVTCADNTGNANNGIVTFGNTSVTVG-----DIGNTNALATVNVGAGTATLGGAVI 454
Qy	481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTTGNKQIQV-G 536
Db	455 KATTKTLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494
Qy	537 ANGKEATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578
Db	495 LNG-ALSOQVTGIGNTSLATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVV---550
Qy	579 APYLDKERLKVGRVEIT--TDSGI--NAGNHKIT---GLTNGIAN-----616
Db	551 -----VTGADNTGNANNGIVTFGNTSVTVGDTGNTSLATISVGAGTATLGGAVI 601
Qy	617 -----TDAVTIKOLKDAKPTL-----NAGDGISINSNNGDLVDSGGNI--TTPYN 660
Db	602 KATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLAT 661
Qy	661 ISVKTTKLNSGTSGNNKFSVSNADHNNSLVTAKDLADYLNKVNETAADSALPSEKVGND 720
Db	662 ISVGA---GTATLGGAVIKATTKITNAVSARK-----FTNPVVVTG-----AIDSTG 706
Qy	721 NSNNAI---TVCKDNTNGKTFNTLKLKGENGVNITTNRATGTVTTFGIDQSNQ-----LTPPK 773
Db	707 NANGIVTFGNTSVTVGDTGIGNTAL-----ATNVGAGTATLG-----GAVIKATTK 754
Qy	774 LTVGSD-----TGNRNLVIEQVPSADGNSTKNIK-----GLSPTLPISIASPSG 817
Db	755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLATISVGAG 814

A:Reference number: A41477; MUID:90354033	
A:Accession: A41477	
A:Molecule type: DNA	
A:Residues: 1-2249 <AND>	
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466	
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi	
C:Keywords: surface antigen; tandem repeat	
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>	
Query Match 5.9%; Score 608; DB 2; Length 2249;	
Best Local Similarity 22.5%; Pred. No. 1.3e-15;	
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;	
Qy	82 INNTLKDALATGEASIFAGSLSKAQGSOAIAIGSVKPPDPPNNGSNVGSN-----132
Db	13 IQOGLKAALFTTSTAIMLSS-----SGALGVATGVIATNNNAFNNYGNWNNWEITAAG 68
Qy	133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176
Db	69 VANGTPAGGPQNMAFTYGGDYTVTTADAADRIKAINVAGT-----TPVCLNI-----116
Qy	177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224
Db	117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFDPADNVTGLGN 168
Qy	225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272
Db	169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223
Qy	273 GRGVALGFGSOILLDRNN-TDASAYVPL-----GKTLADQYKATROGSDTDFSI-----321
Db	224 GAGTATLGGAVIKATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283
Qy	322 -GNSNNNSSTIRKIIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374
Db	284 TGDIGNTS---LATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVVVTCADNTG 340
Qy	375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCGNGLKVLKAKELTGLTSVSAT 427
Db	341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLLQVGVVVKANTINLTNDSAVTET 399
Qy	428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG----480
Db	400 NPVVVTCADNTGNANNGIVTFGNTSVTVG-----DIGNTNALATVNVGAGTATLGGAVI 454
Qy	481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTTGNKQIQV-G 536
Db	455 KATTKTLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494
Qy	537 ANGKEATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578
Db	495 LNG-ALSOQVTGIGNTSLATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVV---550
Qy	579 APYLDKERLKVGRVEIT--TDSGI--NAGNHKIT---GLTNGIAN-----616
Db	551 -----VTGADNTGNANNGIVTFGNTSVTVGDTGNTSLATISVGAGTATLGGAVI 601
Qy	617 -----TDAVTIKOLKDAKPTL-----NAGDGISINSNNGDLVDSGGNI--TTPYN 660
Db	602 KATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLAT 661
Qy	661 ISVKTTKLNSGTSGNNKFSVSNADHNNSLVTAKDLADYLNKVNETAADSALPSEKVGND 720
Db	662 ISVGA---GTATLGGAVIKATTKITNAVSARK-----FTNPVVVTG-----AIDSTG 706
Qy	721 NSNNAI---TVCKDNTNGKTFNTLKLKGENGVNITTNRATGTVTTFGIDQSNQ-----LTPPK 773
Db	707 NANGIVTFGNTSVTVGDTGIGNTAL-----ATNVGAGTATLG-----GAVIKATTK 754
Qy	774 LTVGSD-----TGNRNLVIEQVPSADGNSTKNIK-----GLSPTLPISIASPSG 817
Db	755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLATISVGAG 814

A:Reference number: A41477; MUID:90354033	
A:Accession: A41477	
A:Molecule type: DNA	
A:Residues: 1-2249 <AND>	
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466	
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi	
C:Keywords: surface antigen; tandem repeat	
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>	
Query Match 5.9%; Score 608; DB 2; Length 2249;	
Best Local Similarity 22.5%; Pred. No. 1.3e-15;	
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;	
Qy	82 INNTLKDALATGEASIFAGSLSKAQGSOAIAIGSVKPPDPPNNGSNVGSN-----132
Db	13 IQOGLKAALFTTSTAIMLSS-----SGALGVATGVIATNNNAFNNYGNWNNWEITAAG 68
Qy	133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176
Db	69 VANGTPAGGPQNMAFTYGGDYTVTTADAADRIKAINVAGT-----TPVCLNI-----116
Qy	177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224
Db	117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFDPADNVTGLGN 168
Qy	225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272
Db	169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223
Qy	273 GRGVALGFGSOILLDRNN-TDASAYVPL-----GKTLADQYKATROGSDTDFSI-----321
Db	224 GAGTATLGGAVIKATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283
Qy	322 -GNSNNNSSTIRKIIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374
Db	284 TGDIGNTS---LATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVVVTCADNTG 340
Qy	375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCGNGLKVLKAKELTGLTSVSAT 427
Db	341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLLQVGVVVKANTINLTNDSAVTET 399
Qy	428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG----480
Db	400 NPVVVTCADNTGNANNGIVTFGNTSVTVG-----DIGNTNALATVNVGAGTATLGGAVI 454
Qy	481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTTGNKQIQV-G 536
Db	455 KATTKTLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494
Qy	537 ANGKEATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578
Db	495 LNG-ALSOQVTGIGNTSLATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVV---550
Qy	579 APYLDKERLKVGRVEIT--TDSGI--NAGNHKIT---GLTNGIAN-----616
Db	551 -----VTGADNTGNANNGIVTFGNTSVTVGDTGNTSLATISVGAGTATLGGAVI 601
Qy	617 -----TDAVTIKOLKDAKPTL-----NAGDGISINSNNGDLVDSGGNI--TTPYN 660
Db	602 KATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLAT 661
Qy	661 ISVKTTKLNSGTSGNNKFSVSNADHNNSLVTAKDLADYLNKVNETAADSALPSEKVGND 720
Db	662 ISVGA---GTATLGGAVIKATTKITNAVSARK-----FTNPVVVTG-----AIDSTG 706
Qy	721 NSNNAI---TVCKDNTNGKTFNTLKLKGENGVNITTNRATGTVTTFGIDQSNQ-----LTPPK 773
Db	707 NANGIVTFGNTSVTVGDTGIGNTAL-----ATNVGAGTATLG-----GAVIKATTK 754
Qy	774 LTVGSD-----TGNRNLVIEQVPSADGNSTKNIK-----GLSPTLPISIASPSG 817
Db	755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQVTDGIGNTSLATISVGAG 814

A:Reference number: A41477; MUID:90354033	
A:Accession: A41477	
A:Molecule type: DNA	
A:Residues: 1-2249 <AND>	
A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA26380.1; PID:g152466	
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi	
C:Keywords: surface antigen; tandem repeat	
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>	
Query Match 5.9%; Score 608; DB 2; Length 2249;	
Best Local Similarity 22.5%; Pred. No. 1.3e-15;	
Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;	
Qy	82 INNTLKDALATGEASIFAGSLSKAQGSOAIAIGSVKPPDPPNNGSNVGSN-----132
Db	13 IQOGLKAALFTTSTAIMLSS-----SGALGVATGVIATNNNAFNNYGNWNNWEITAAG 68
Qy	133 -----AKNVESTAIGDVLAEGBASIAI-----GSDDLVLPKNLDLKNFEHKLII 176
Db	69 VANGTPAGGPQNMAFTYGGDYTVTTADAADRIKAINVAGT-----TPVCLNI-----116
Qy	177 HGHEILKKIQTSTDGKIKYRR-----TRAQGHASTAVGAMSYAQGHFSNA-----FGT 224
Db	117 -----TQNTVVGSIITKGNLLPVTLNAGKSLTLNGNNAVAANHGFDPADNVTGLGN 168
Qy	225 YATAEAAYSLAVGLAQATKQSSIAVGSN-----AKANAFATAAIGNT-----VNL 272
Db	169 IALGGANAALII-----QSAAPSKITLAGNIDGGIITVKTDA-AINGTIGNTNALATVNV 223
Qy	273 GRGVALGFGSOILLDRNN-TDASAYVPL-----GKTLADQYKATROGSDTDFSI-----321
Db	224 GAGTATLGGAVIKATTKTLTNAASVLTLTNANAVLTGAIDNTTGGDNVGVNLNGLSQQV 283
Qy	322 -GNSNNNSSTIRKIIINVAGSRDTD--AVNVAQLKLVLELANRKIT-----PKGDGDNNS 374
Db	284 TGDIGNTS---LATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVVVTCADNTG 340
Qy	375 NSVERGL---GNLTITKGAQTNALTEANIG---VVTDCGNGLKVLKAKELTGLTSVSAT 427
Db	341 NA-NNGIVTFGNTSVTVGNVGNATLATVNVGAGLLQVGVVVKANTINLTNDSAVTET 399
Qy	428 NKITVNTNNNAELQSGGLTF---SPITGKTDKTVYSIDGLKFTNDSNSIATKG----480
Db	400 NPVVVTCADNTGNANNGIVTFGNTSVTVG-----DIGNTNALATVNVGAGTATLGGAVI 454
Qy	481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNNTTGNKQIQV-G 536
Db	455 KATTKTLTNAASVLTLTN-----ANAVLTG---AIDNTTGGDNVGVNLN 494
Qy	537 ANGKEATVANNVANT-----SATVG-----TARITEEKIGFAGTNDGVDEQ 578
Db	495 LNG-ALSOQVTGIGNTSLATISVGAGTATLGGAVIKATTKTLTDAASAVKFTNPVV---550
Qy	579 APYLDKERLKVGRVEIT--TDSGI--NAGNHKIT---GLTNGIAN-----616
Db	551 -----VTGADNTGNANNGIVTFGNTSVTVG

QY 818 RNTALGNTIEKDK--SNAASIDVVLNAGFNKNNKDKDFYSTVDTVFIDGNATTATV 875
Db 815 TATLGAIVKATTTKLTNAASVLTNANAVLTG---AVDNTTGGDNVGVNLNGLSQQV 871
QY 876 TYDEANQTSKVAVDVNDEKTIETLGDNGKKQKGVKTIKLE-----TSNGNATTFSTD 930
Db 872 TGDIGNTSLATISVGAGTATL-----GGAVIKATTTKLTNAASVLTNANAVLTGAI 925
QY 931 D-----DHALVKASDIAGNLTALAEIHTTKGTA---NTALQTFVKKVDEN 974
Db 926 DNTTGGDNVGVNLNGLSQTGIDGNTSLA--TISVGAGTATLGAIVKATTTKLTDA 984
QY 975 DKADDNATV--GKDGTSKVNLTCLKKNGLDIKDKDGTVTFTGNTOSGLKAGDSTTL 1033
Db 985 SAVKFTNPVVVTGAIDNTGNAN-----NGIVTF-----TGNSTVT 1019
QY 1034 NNGLSIIKNTASNEQIOVGADGVKFAWNVGVGAGIDGTTTRITRDEIGTGTN-----G 1088
Db 1020 GNVG-----NTNALTAVNVGAG---LLOVGGVVKA---NTINLTDNASAVFTNPVVVTG 1069
QY 1089 SLDKSPHLSKDGINAGGKKITNIQSGEIAKNSHDAVTTGGKIYDLKTELENKISSTAKTA 1148
Db 1070 AID-----NTG-----NANGIYVFTTGNSTVTG-----NVGNT----- 1097
QY 1149 QNSLHEFSVA-----DEQGNFTVSNPYSSYDTSKTSVITFAGENGITTKVKNGVVVRGI 1204
Db 1098 -NALTAVNVGAGLLOVGGVVKA---NTINLTDNASAVTF-----TNPVVVTGAI 1143
QY 1205 DQPKGLTTPKLTGVN--NNGKGIWINSQONTITG--LSNTLANVNTDKGSVRTEOGNII 1262
Db 1144 DNT-----GNAING-----IVFTTGNSTVTGIDGNTNALTAVNVGAGITLQAGSL 1189
QY 1263 KDEDKTRAASIVD-----VLSAGFNLOQNGEAVDFYSTVDTVNFANGNT----- 1306
Db 1190 -----AANNIDFGARSTLEFNGPLDGGKAIPY---YFKGAIANGNAILNVNTKLT 1239
QY 1307 -----TTAKVYDDTSKTSKVVDVNVDDTTI-----EVKKKLGKVTTLTSTGTGA 1354
Db 1240 ASHLTGTVAEINIGAGNLTIDASVGDVITLNAQINFRARSDSVLVSML-----TCGVV 1295
QY 1355 NKFPALNQATGDALVASIDIVAHNLTSGLDIQT-----AKGASOANNSAGYV---DAD 1404
Db 1296 NNTLL-----AADLVAP-----GADEGTVPVNGVNGVNLVNSVAGTARNITGDGG 1340
QY 1405 GNK-----VIYDS-----TDNKYYQAKNDGTVDKTKEVAK-----DKLVA 1439
Db 1341 GNKFTLLIYNAVITDDVNLGIONVLINKNADFTSSAFNAGAIQINDATVYIDANNG 1400
QY 1440 QAOTPDGTL--AOMNVKSVINKEQVNDANKKQGINEDNAPVKGLEKAAADNKTNAATV 1497
Db 1401 NLMIPAGNIQFAHADAQLVLQSSGNDRTITLGANTID-----PDNDDEGIVI-- 1447
QY 1498 GDNAV-AQTPLTFAGDTGT---AKKLGELTIKG--GQDNTNKLTDNNIGVAGTGDFT 1552
Db 1448 --LNSVTACKKLLTIAG--GKTFGGAHL--OTILFKGAGDCST-----AGTFTNT 1491
QY 1553 VKLAKDLTNLNSVAGGTKIDERGISFVDANGQAKANTPVLSANGLDLGGKVISNVGKGT 1612
Db 1492 TNIVLDIT-----GQ-----LELG----- 1505
QY 1613 KOTDAANV-----QQLNEVRNLLGLGNAD-----GNQVNIADIKKDPNKGSSNRT 1660
Db 1506 --ATTANVLFNDVAVLQUTGTNTGGFLDFPNKNGMVTLLANNVNVAGAVQ--NTGGTNGT 1561
QY 1661 VIKAGTVLGGKGNDEKLA---TGGVQGVCKDGNAN--GDLSNVVVKTKQKDSKALLA 1716
Db 1562 LI-----VLGASNLNRVNGIAMLKVAGNVTIAGKKVKIGEIOGTGNT-----ITLPA 1611
QY 1717 TYNAAGQNTVVTNPNPAEADIRINEQIGIRFHVNDGNOEPVQVRNGIDSSASCKHSAIG 1776
Db 1612 HFNLTG-----SINKTGGQALKLNFNMNGSGVGV----- 1641

QY 1777 FQAKAGEAAVATGROTAQAGNOIAIGDNAQAT-----GDOSIA---IGTGNVWAGKHSGA 1829
Db 1642 -----GTAANSVCDITTAGATSFASSVNAKGTATLGGTTFANTFTNTGAVTLAK----- 1691
QY 1830 IGPSTVVKADNSVSVGNNOFTDATQTDVFGVGNNTVVTESNVVALGSNAISAGTHAGT 1889
Db 1692 -----GSITSFAKNVTATSF--VANSATINFSNLSAFNSN--ITGG----- 1728
QY 1890 QARKSDGTAGTTTTTGA-----TCTVKGFAGQTAAGVAVSVGA--SGAERRIQNVAAEVS 1942
Db 1729 -----GTTLTLCANQVYTGTT--GSFTDTLTNTTFDGAAGSGGNILIKSGTSLDLS 1778
QY 1943 ATSTDVAVGSQLYKATQSIANATN 1966
Db 1779 GVSTLAL-----VVTATN 1791
RESULT 6
B71704
cell surface antigen (sca3) RP451 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: B71704
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: B71704
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2340 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAAL4908.1; PID:g3886
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: sca3; RP451
Query Match 5.7%; Score 591.5; DB 2; Length 2340;
Best Local Similarity 21.5%; Pred. No. 5.9e-15;
Matches 534; Conservative 307; Mismatches 839; Indels 805; Gaps 139;
QY 35 TGOVGSVRTLSFARI---AALAVLVIGATLNGSAYAQIITKIEIGOTNKINNTLKGDAL 91
Db 123 TSIITGVGTGTTFNIDFAGKNATLQINKDLN-----ITTKID-----NTVAG--- 164
QY 92 ATGEASIAFGSLSKAQSQAIGSVKPDPNNGS-----NGNVGSHAKGNESIAIGDVL 146
Db 165 -----YAPEANNITINAKNINLTNNLSL---TLCDGNI----- 230
QY 147 AEGDASIAIGSDDLYLKP--NLDLKNFHLIHHGHEILKKIQTSTDGKIKYRRTAOGHA 204
Db 193 GNCEAKI-----YAPEANNITINAKNINLTNNLSL---TLCDGNI----- 230
QY 205 STAVGAMSYAQGHFSNAGFYATAEAYSLAVGLAAQATKQSSIA-VGSNAKANAFATA 263
Db 231 -----TTLKGNIINT--TEIDGQGLNLAYDLGSSIIITGIDIGTGLDITINVLGSA 281
QY 264 IGGNTVTVNLGRVAGFGSGOILDRDNNTDASAVVPLGKTLADQVKATROGSDTDFISGN 323
Db 282 TFSNTILK-ATNINLKHNTSTLNDN-----IIVIGN 313
QY 324 SNNNSNSIRKIIN--VGAGSRDQDAVNVAAQLKVEELANRKITTFKGDGDNNSN----- 375
Db 314 IKGNNN---KDILNFKVHGTNLDNEMIIPAPQK-----TGTNLNFKGNATLNGNINLNI 365
QY 376 -SVERGLGNTLTITKGDQATNALTEANIGVVTG-----NGL-----KVKLAKELTGLTSV 424
Db 366 LKFSGGHGKTLNFGQTKVDNLVFAD--SVLDSGTISVNGLLDTCVTFNNSNVNGGFLII 424
QY 425 SATNKITVSNTNNNNAELQ--SGGLTFES--PITGKT-----DKTVYSTD-----GLKFTNDS 473
Db 425 NAKNTSAKULLNATKAKIQIKNANLTMNHPSGAGDISDRIADNTIYTIIDAKNGNVNLLNN 484

QY 474 NSIATKG-----TTRIKKK-----IGFAGTND-GV-----DESKPY 504
 Db 485 AKIIFEGADSMALINTGVADRTFTTYNNLNSGNDYGVKIVKAEKKAIVILANOSGPY 544
 QY 505 L---DN-EKLA-----VGNSTLNSGLRVNNT---TGNKQIOVGANG-IRF-----542
 Db 545 TIGQDNTIRKELIVEGAGDIIIDTFTKLLSINSTGQITENRTLDLGGAGNIAFGKIH 604
 QY 543 ATVANNVANTS-----TVCARITEKIGFAGTNCVDEQAPYLDKERLKVGRVE 593
 Db 605 TLVVGVTGSIITSENNOGILITNSGNIT---GVIGTNE-----LGLKLVNIGADP 652
 QY 594 ITTDSGINAGNHKIGTLN-----GIANTDAVT-----IKQ 624
 Db 653 VTCSANVFAS-----VALNPPSVLILADGVTLTGEVTHNNTKGVLSIGTSGNITGQIGT 708
 QY 625 LKDAKPTLNAGDIS-INSN-----NGDLVDSGNTTTP-----657
 Db 709 NSAALEKINIGAGASNIDSNIVAGSTVLTDQTSLETLLNNDVVVNSNIITAGNNSGKLIF 768
 QY 658 -----TYNISVKTTL-----NSNGTSGNNKFSVNSAHDN---NSLVTA-----KD 695
 Db 769 TNGGIGTNGIGANGAALQEWFFNGTTCGTPANSQNFVVAHSAAVNVITGLTTGALKYKD 828
 QY 696 LA-----DYLKNE-----TADSLPSEKVO-NG-----DNSNNAITVGKO 731
 Db 829 TGTIIAHGVLVDIFENNRKAGKFIKLGAMIDGVLGNGVAGTLDLFGDGNVTGONIGAD 888
 QY 732 TNGKTFNFKLKGENVNITNATGVTFVFIQDSNG-----LTPPKLVGSDTNGN 783
 Db 889 -NANSISTINQDNTKNVTI---ANDIFVDNIHTNGGILQGLNLTTHNIDFGA---NGG 943
 QY 784 RLVTIEQVPSADGNTKNIKGLSPTLPSTASPSGRNIALGNTIEEKDKNASIDDLVNA 843
 Db 944 TL-----EFGNNTYLN-----NALIVGQNGILNAFTNLKASODTGTGVKILNI 988
 QY 844 G-----FNLKNGKDKDFVSTYD-TVDFIDGNA-----TTATVTV-DEANQTSKYA 887
 Db 989 GOIGTPPOFTTOVNNKNLTLVSSVNSINFGDANSQILLSAPVDQTLKFNNLNETGGGI 1048
 QY 888 YDNNVDEKTIETLGDNGKQKLGKVTIKLITETSTNGNAT-TFSTYD-----D 931
 Db 1049 ITLDSGNNLTISGNGIK-LGSKNELSSLNKGKVTYVNDLIDQIHQNLNNGALFD 1107
 QY 932 DHALVKASDIAGNLTLAEETHHTKGTANTALQTFVTKVDENDKADDTNATVTKGDGTS 991
 Db 1108 DOSLTSK---IKNIN-----IGTVAGGA-----TYLDAINDN---FDLN-----TS 1144
 QY 992 GKV-----NTLKLKGNGLDIKTDKGVTFGI---NTQSG-----LKAGDSTLLNNGLS 1039
 Db 1145 GWVFKHQSILELKNSSNTN---DHTITLTSALDPGNNQFQIILITDTNKLFDING-- 1199
 QY 1040 IKNTA-----SNEQIQVGADGVKFAWVNGV-----VGAGIDGTRITRD-BIGFTGTNGS 1089
 Db 1200 --NVAYTLGTANMLK-----QLTFASIDGALAKAVGINNVNVLNLIKDIELNEVNAVNL 1253
 QY 1090 LDKSKPHLSKDGINA-----GKKITNIO-----SGEAKNSHDAVTC-----GKIYD 1132
 Db 1254 FNKNVTVTATGNIHGVDFQGNAGVINLNDIDIEIDGVSITSGN---VNGTLNFGSGKVTG 1311
 QY 1133 LKTELENKISSTAKTQNSLHEFSVADEQG---NNFT-VSNPYSSYDTSKT-----1179
 Db 1312 LINNIVMLQAGADVLSASGNSYIETIOGNGNNLTFAANSHLTDTINKTGQDNLNVF 1371
 QY 1180 -----SDVTTFAG-----ENGIITTKVKNKVVVRVGDID-QT 1207
 Db 1372 INGSVSGSIGANAAGVDIIINAGSNNFNSLTKSGNIVISDGTQVMVNNVTATDISGN 1431
 QY 1208 KGLTTPKL-----TVNNGKGINVNSQNGTIT-----1237
 Db 1432 ANNGTLKLNHTPNTITSTLGNNAIG-TIEVANNNDVTITGLTQAOQNIHFSNATQAATLT 1490
 QY 1238 -GLSNTLANVTNDKGSVKTTE-----OGN---IIDEKDKTRAASIVDVLVSAGFNLOQNG-- 1287

Db 1491 LGAASQVNTTITAGNIHTLEVTFDTGNDGIIGDAN-NRLKSI-----ELTGNQTV 1541
 QY 1288 --EAVDFSVTYDVTNFGANG----TTTAKVTYDDTSKTSKVVDVNVDDPTI---EYKDKKL 1340
 Db 1542 TINSPHYSSITTANNAQGNVKLINEGGIYDGLSKISLANVQISEDTTIRGDVYSKYL 1601
 QY 1341 GVKTTTLTSTGTGANKFALSNQATGDALV-----KASDIVAILNTLSD-IOTAKGASO 1393
 Db 1602 NIDAGTIINDRGNNMKNLNDIPDALIDVLPRLSLFNFYTDIKADNLNFADDTAT 1661
 QY 1394 AN-NSAGVYVDAD-GNKVIYDSTONKY---YQAKNDGTVDKTKVEAKDKLVAQAQTPDGTILA 1449
 Db 1662 ANFKDAVVIDAHIDNGGILKFNFDNWLTOEIKNANII---EIASDKFM-----LL 1708
 QY 1450 OMNVKS-----VINKEQVNDANKKOGINEDNAFVKLEKAASDNKTKNAAVTVG 1498
 Db 1709 QKNIKAAATLIADNANLVLDDNVEVNTNLNVRDILVLDLA---NYELKYTGNTINGLTLTI- 1764
 QY 1499 DLNAVAQTPLETFAGDTGTAKKLGETLTIKGSQTDTKNLTDNNIGVVAGTDGFTVKLAKD 1558
 Db 1765 -----ITY---FDYALQKGGHILVSQGSNVDMSDL-DNLI-----IKIKAHSD 1803
 QY 1559 LTNLNS-----VNAGGTKIDKIGISFVDANGQAKANTPVLNANGL-----1598
 Db 1804 ITNITSDTKHQIVKLETGAIYTPVQTKVIIDASEEQNKFKVKKVADANGLVLLTDTGGRD 1863
 QY 1599 DLGGKVIS-----NVGKGTQDT-DAANV---QOLNEVRNLLGLCNDNADGNVNIADIRK-- 1649
 Db 1864 DTGGRDTRGRGTNGCRDNCVGNISNSSNEA-----GGSSDDKN-YGITDVPPIF 1916
 QY 1650 DPN-----SGSSS---NRTVIRKAGTVLGGKGNNDTEKATGCVQGVQDKDG-- 1692
 Db 1917 DPSILDYTKNNYVAGSLANLHVHKDFGNTTDACKLLNDLGFMSPNRVTTETDLRLSNR 1976
 QY 1693 -NANG-----DLSNVWVK-----TQKD-GSKKALLA---TYNAGQNTVYT 1728
 Db 1977 INVNGLNEGVLNGIEVENFLTDIAINMDFNTAKEIGHRLBELSDANTVINGLNTLTL 2036
 QY 1729 NNPRAEADRINEQGIREFHVNDGNQEPVQVGRNGIDSSASGKHSVAIGFOAKADGAAVA 1788
 Db 2037 NKK-----INLKRNTNQAIIAAGDE---DNIVTGIWMSFYGKIKONSKNS-A 2082
 QY 1789 IGROTAQNSIAIGDNNQATGDSIAIGTGVVAG---KHSGAIGDPPSTVKADNSYSVG 1845
 Db 2083 SCYOSNTGGGIIIGFDYNI---DNSIVIGAAVYMADESKVHKHKNKNGDRUKAKSNISYI 2138
 QY 1846 -----NNQFTDA-----TOTDFVGVNNTVITESVALGSNAIS 1882
 Db 2139 GLYNLNTNFFVEAIGVYGRNKKIKNYEKRIITITDQIAGKFINTFYSYELLGYNLIS 2198
 QY 1883 AGTHAGT-----QAQKSDGTAGTTTAGATCTVKGFAG--QTAVGAVSVGASGAER 1931
 Db 2199 ---HRTTIPMFGMYATFKNNYKENNTTQNLISIKKNYVDKFTILGLNSV-----2248
 QY 1932 RIONVAAGEVSATSDAVNGSQL-----YKATQSIANTATNELDHRIHONENKANAGISSA 1986
 Db 2249 -----THYLSQDILIKPELHWFVINYQCKNKLPNIDARLD-----GIDEP 2287
 QY 1987 MAMASMPQAIYIPGRSMVTGCIATHN 2011
 Db 2288 LTTIRFKPAKITYN--LGGGISTKN 2310

RESULT 7

F81045
 hemagglutinin/hemolysin-related protein NMB1768 [imported] - Nelisseria meningitidis (C)
 C:Species: Nelisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: F81045
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. F.I., H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vezzani, A.; Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: F81045
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2514 <TET>
A:Cross-references: GB:AE002326; GB:AE002098; NID:g7227015; PIDN:AAF42109.1; PID:g722702
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1768

Query Match 5.3%; Score 553; DB 2; Length 2514;
Best Local Similarity 20.0%; Pred. No. 1.9e-13;
Matches 503; Conservative 332; Mismatches 895; Indels 782; Gaps 121;

QY 4 IYKVFENKATGTFMAVEYAKSHSTGGSCATQGVGS--VRTLSFARIAALAVLVIGATL 61
DB 5 LYRVIFENKRGAVVAETTKRE---GKSCADSDSGSAHVKSVPFG----- 47

QY 62 NGSAYAAQITTKTEIGQTKINNTLAKGDA--LATGEASTAFG-----SLSKAQGSQAI 112
DB 48 -----TTHAPVCRSNIFSFSLGSLCLAVGTANTAFADGIIADKAAKPTQATIL 98

QY 113 AIGSVKPDN-----NCSNGNVGSHAKGNESIAIGGDV-----LAEG 149
DB 99 QTNGGIPQVNIQPTTSAGVSNOYAFQDVGNGKAILNNSRSNTQTQLGWIOGNPWLARG 158

QY 150 DASIAIGSDDLIYLPKNDLKNFEHLKHGHEILKKTQTDGKIKYRTPAAGHASTAVG 209
DB 159 EAAVWV-----NQINSHSSOMNGVIEVGGRAEVVIANPAG 195

QY 210 AMSYAOCHFSNAGTYATAEAYSLAVLAAQATQSSIAV---GSKANANAF----- 259
DB 196 IAVNGGFFINASRATLTGPOQYQ-AGDLGFKIROGVNVIAGHGLDARDTDFRILSYH 254

QY 260 -----AATATGGNTVNVMLRGVALGFGSQILDRDNT-DASAVPL-----GK- 301
DB 255 SKIDAPWGDVVRVAGQNDVATGNAHSPILNNAANTSNNTANNGTHIPLFAIDTGKL 314

QY 302 -----TLADQYKATRGDSTDIFFSIGNNNNSIRKILNVG---AGSRDID 346
DB 315 GMYANKITLIATAEQAGTRNOG---OLFA--SSGNAVIDANGRLVNSGTMAAANAKDID 369

QY 347 AVNVAQLKVEELANRKITFKPGDGNNSNSVERGLGNTLTIKGAQTNALTEANIGVVTD 406
DB 370 -----NTAEHKVNIRSGVENS----- 387

QY 407 GNLKVKVLAKELTGLTSVATNKITVNTNNNAELQSGGLTFSPITGKTDTKTVYSIDG 466
DB 388 -----TAVSQOQTQIHSQSIQNTGTLSSGEILIHNSGSLKNETSGTIEA 432

QY 467 LKFTDNSNATKG---TTRITKKKIGFAGTND-----GVDESKPYLDNEKLKVGNSTLN 518
DB 433 ARLAIDTDLTNNQKLSQSGSKLHDAQCKMDNRMGLQDAPTASNGS-----SNQTN 489

QY 519 SGSLTVNNTTGNKIQIVGANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDEQ 578
DB 490 SYNASEHSSTTTTATG-TGTATVSIISNITAPTAD-GTIR-----THGALDNS 537

QY 579 APYLDKERLKVGRVETTDGSI-NAGNHKITGLT-----NGIANTDAVITKOLKDA 628
DB 538 GSIIAN-----GQTDVSAQOGLNAGQIDHQLNAKGSAPDNHNGTILISDAVHIQ----- 587

QY 629 KPTLNAGDGISNSNGDLVDSSGNTITPTTYNISVKTTL----- 668
DB 588 -----AG--SLANQN-----GNITT-RQQLIEITDQLDQNAHGLLSAETIADLAVSGS 631

QY 669 --NSNCTSGNKFVSNAHD-----NNSLVTAKDLA---DYLNKVNETAADSL 711
DB 632 LNNQNGEIAITNQOLI--IHDGQOSTAVIDNTNGTIQSGRQVIAQKSLNSNGTLAANDKL 689

QY 712 -----PSFKVQNGDGNNSNATVCKDNGKTFNTLKLKNGVNIITNRTATGTVTFGIDQS 766
DB 690 DIALQDDFYVERNIVAGNELSL--STRGSLKNSHTLOAGKRIRIKANNLNDAAQGN-QS 746

QY 767 NGLTTPKLTVDSTN-GNRLVIEQVPSADGNSTK-----NIKGLSPILPSTASPSGR 818
DB 747 GGT-----DIGTQHNLTNRGL-----DQOQTKIOAGQNNIGTG-----RIYD 787

QY 819 NIALGNT-IEEKDKN---AASIDDDVLNAGFLNKNNGKDKDFVSTYDVFIDGNATTAT 874
DB 788 NIAIAATRLDQDENGTAIAAARENLNIGQLNNRENSLIYS-----GNDMAVG 838

QY 875 VYDEANQTSKVAYDVNVDEKTIETLGDNGKQKQGVKTI-----KITETSTNGAT 925
DB 839 GALTDTNGQATGKAQRHAGATIEAA---GKMRLGVEKHLNTHNEHLTKQLVETGRE-HIV 894

QY 926 TPESTDDHVALVKAS---DIAGNL-NTLAEIHTTKGTANTALQTFVTKKYVDENDKADDT- 980
DB 895 DYEAFGRHELLREGTOHELGSVYNDESHLRTPDGAHENWHKYDEKYTKQTKVTQTA 954

QY 981 -----NALTV-GKD--GTSKGVNTLKLKNGKGLDITKDKG---TVTFG---INTQSG 1024
DB 955 PAKIISNDLTDIGKEVFENDSQI-----IAGGN-LIVQTEKDLHNEQTFGEKKVPSENG 1009

QY 1025 L-----KAGDST-----TL-----N 1034
DB 1010 KLHSYWRKHKGRDSTCHSBNQVNTLPEEITRNISLGSFAYESHKALSHHAPSQGTLPQ 1069

QY 1035 NGLSIKNTASNOIQVADGVKFAVMNVNGVAGIDGTTTIT--RDEIGFTGTNGSLDK 1092
DB 1070 SNISLPTYSNTPTLPSLSLYIINPVNKYL---VETDPFANYRQWLGSYDMLDSLKL 1126

QY 1093 SKPHLSK---DGINAGKKITNQSGEIAKNSHDAVTGGKIYDLKTELEKNISAKTAAQ 1149
DB 1127 DPNNLHKRLGDIYVE-QRLNEQIAEL--TGHRLDQ---YONDEQFALMDNGATAA 1179

QY 1150 NSLHESVADDEQNNFTVSNPYSDYTSK-TSDVITFAGENGITTKVNGVVRVIGIDOTK 1208
DB 1180 RSM-----NLSVGIALSAEQVAQLTSDIVMLVQKE-----VKLPDGGTQ 1218

QY 1209 GLTTPKLTV---GNNGKGVINSQNGNTITG-----LSNTLA 1244
DB 1219 TVLPVQVYVRKNGIDGKGLLSGNTQINVSGLNSGTIAGRNALIINTDPLDNG 1278

QY 1245 NVTFNDKGSVFTTQ---GNIIKDEKTRAASIIVDLSAGFNQNGEAVDFVSTYDVTN 1300
DB 1279 RIHAQKSAVTATQDINNIGGMLSAEQTL-----LLNAGNNI----- 1314

QY 1301 FANGNTTAKVYDDTSKTSKVYDVNVDDTTIEVKDKLGVKTTTLTGTGTANKFALS 1360
DB 1315 --NSQSTTASSQNTQSGST-----YLDRMAGIYI-----TGKEKGVA 1350

QY 1361 NOATGDALVKASDIVAHLNLTSLGDIQTAKGASQANNAG-----YVDAD 1404
DB 1351 AQAKD-----INIIAGIISQSGOQTRLOAGRDINDTVOTSKHOATHEDAD 1399

QY 1405 GNRKVIDSTD--NKYQAKNDGTVDKTEKAKDKLVAQAQTPDGTLA-----QMNKSVI 1457
DB 1400 -NHVIRGSTNEVGSSIQTKGDVTLSSGNNL--NAKAAEVSSANGTAVSAKNDINISAGI 1456

QY 1458 NKEOVNDANKQGINEDNAFVKGLEKAAASDNKTKNAAVTVG-----DLNVAQTP 1508
DB 1457 NTHVDDASKHTGRSGGKGLVITDKAQSHHETAQSSFFEGKQVVLQAGNDANILGSNV 1516

QY 1509 TTAGD-----TGTTAK-----LGETLTKGGOTDNKL 1537
DB 1517 SDNGTQIAGNHVRIGTQTQSOSETYHQTKSGLMSAGIGFTIGSKTNTQENQSOSEH 1576

QY 1538 TDNNIGVAGTGTGVKLAIDLNLNSVNA---GGTKIDEKISFVDANGOAKATP-...1591
DB 1577 TGSTVSLKGD-TIVAGKHVEIGTGVSSPEGNNWIIYAQSIDIOAHKLNLSNTQTY 1634

Thu Sep 13 14:19:27 2001

```
QY 1592 -----VLSANGDLGGKVIS-----NYGKGTGK-----TDAANY--OOLNEVRNLLGL 1632
Db 1635 EQKGLTVAFSPVTDLAQQAIAVAQSKQVQSKNDNRVMAAANAGWQAYQCKSAQNL 1694
QY 1633 GNDNADGNOVTADIKKDPNPGSSSNR--TVIRKAGTV-----LGKGNNDTEKLATGGV 1684
Db 1695 ANGTTNAKQVSI-----ITYGEQNRQTTQVQANQAQASQIQAGGK-----TTLIATGAA 1745
QY 1685 Q-----VGVKDQGNA-----NGDLSNVVVKTKQSGKKA-----LLATYNA 1720
Db 1746 EQSNINAGSDVAGKAGTILADNDITLQSAEQSN-----TERGONKSAGNAGAAVSFGQ 1801
QY 1721 AGQTNVYT--NNPAAIDRINEQIRFHVNDGNOEPVVOGRNGIDSSASGKHSVAIGFQ 1778
Db 1802 GWSLGVTVAGGNGVGYGNGDSITHRSHIGDKGSQTLIQ--SGGDTTIRKAQVGRGKVQ 1859
QY 1779 AKA-----DGAAVAIGRQTOAG--NQSIAIGNAQATQDSIAI 1816
Db 1860 VNAKNLSIQSDRETYQSKQONASAOQTVGVGSAGGDYSQSKIRADHVSVTEQSGIYA 1919
QY 1817 GTG--NVVAGKHSAGLD--PSTVRADNSYSVGNNNQTFDATOT--DYEGVGNNTVTES 1870
Db 1920 GEDGYQIKVGNHHTDLKGIIITSTQSAEDK---GKNRQFATLTHSDI---KNHSQYKG 1971
QY 1871 NSVALGSNSAISAGT--HAGTQAKKSDGTAGTTTAGATGTVKGFAGOTAVGAVSVGASG- 1928
Db 1972 ESFCLGASASISGKTLGQGAQNKPNKHLTSVADKNSASSVGY--GSDSDSQSSITKSGI 2030
QY 1929 -----AERRIQNVAAAGSVSATSDAVNGSOLYKATQSIANATN-----ELD-- 1969
Db 2031 NTRNIOTDRAAQIRLGTAKTADIDNTVITDPAERHSGSLKNTFNKEAVQSELDIQ 2090
QY 1970 ----HRIHQENKANAGISSAM-----AMASPOAYIPGRSVMTGGIATHNGOGAV 2016
Db 2091 RTVSQDSFKNVQQAQNTIENHQLKADKADPAETAATAAALANG--DMETAKRAHEAQDAA 2149
QY 2017 AVGLSKLSDN---GOWFKINGS---ADTOGHVGAAGA-----GFHF 2053
Db 2150 A-----KADNMQQQGVILNMLASGLAAPTQSGAGIAATAASPASVVAIGQHF 2196

RESULT 8
E85822
probable invasins 23135 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <STO>
A:Cross-references: GB:AE005174; NID:g12516151; PIDN:AAG57041.1; GSPDB:GN00145; UWGP:231
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 23135

Query Match 5.2%; Score 542.5; DB 2; Length 2660;
Best Local Similarity 20.9%; Pred. No. 5.1e-13;
Matches 551; Conservative 313; Mismatches 934; Indels 837; Gaps 132;

QY 47 ARIALAVLVIGATLNGSAYAAQITTK--TEIGQTNKIN-----NTLKGDAL--- 91
Db 59 AQIAIANANTVPYTLGALESQAQSVABRFGISVAELRKLNQFRTFARGPDNVROGDELDP 118
QY 92 -----ATGASIAFAGSLSKAQSQAQIAIGSVKPPNNGSNGVNGSHAKGNESIAIGQ- 143
Db 119 AQVSENNLTTPPGNSSGNLEQIATSTSQIGSLIAEDMNSE--QAANMARGWASSQAAGA 176
```

```
QY 144 --DVLAE--GDASIAIGSDLLYLPKNLKLKNEFHKLHG-----HEILKKI 185
Db 177 MTDLRSRFGTARITLGVDE-----DPSLKNSQDFDLHPWYETPONLFFSQHTLHRTDERT 231
QY 186 QTSTD--GRIKYRRTRAQ-----HASTAVGAMSQAQGHF--SNAGFTVATAE 229
Db 232 QININGLWRHFTPTMWSGINFDFHDLRSYHSRAGIGA--EYWRDYLKLSNGYLRLLTNWR 290
QY 230 AAYSLAVGLAAQATKOSSIAVGSNAKANAF--AATAIGGNTVVN--LGRGVALGFGSQIL 285
Db 291 SAPELDNDYEARP-----ANGWVRAGWLPANPHLGGKLVEYEQYDEVAL--FDRD-- 341
QY 286 DRDNNTDASA-----YVPLG--KTADQYKATQGSTDFISI-----GNS-----NNNS 329
Db 342 DQSNPHAITAGLNVTPPLMTFSAEQRQKGQENDTRFAVDFTWQPGSAMQKQIDNEV 401
QY 330 STARKIINVGASRQTDV-----NVAQLKVLVEELNKRKITFKGDDNNS--NSVERGLGNT 384
Db 402 DARRSL-----AGSR--FDLVDRNNNIVLYRKKELVRLTIDPVTGKGEVKSLSVLTQTK 456
QY 385 LTIKG--DAOTNALTEANIGVITDGNGLKVYL--AKEITGLTSVSAT--NKIIVSNNTNNA 440
Db 457 YALGYNVEATALEAAGGKVTTGKIDILVTPAYRFTSTPETDNTWPIEVTAEDVKGNFS 516
QY 441 ELQSGGLTSPITGTDKTVYSIDGLKFTDNSNSTATKGTTRITKKIG-----FAG 493
Db 517 NREQSMVWVQAPLTSOKDSSV--SLSQTLISAQSHSAT--LTFIAHDAAGNPVIGLVLST 573
QY 494 TNDGVDE--SKPYLDNEKLKVGNSSTLNSGS---LTVNNTGNKQIOVGANGI---KPAT 544
Db 574 RHEGVQDITLSDKDN-----GDGSYQILTGTGAMSGTLTLPOLNGVDAAKAPA 623
QY 545 VANVA--NTSATVGTARITEEK-----ICFAGTND--GVDEQAPYLDK-----ERLK 588
Db 624 VVNTISVSSRTHSSIIKDKDRGLSNPPIEVTVELRDENDKPKVEKQKQOLNTAVSIDNVK 683
QY 589 VGRVEITD-----SGINAGNH-----KITGLTNGIA-----NTDAVTIKOLKDAKP----- 630
Db 684 PG---VITDMKETADGVVKATYATYTKGSLGTAKLLMONWNEDLHTAGFIIDANOSAKI 740
QY 631 --TLNA--GDGSIINSNGDLVDSSGNI---TTPTYNISVKTTLKLSNGTSGNKK----- 678
Db 741 ATLSASNGVLIANEAANTV--SVNVADEGSNPINDHTVTFVAVLSGSAFTFNQNTAKTD 798
QY 679 -----FSVSNADHNNLSVTAKDLADYLNKVNEDASALPSEKQVQNGSNNALTVGK 730
Db 799 VNGLATFDLKSSQKQEDNTEVTELENGVVKQTLIVSFVGDSSSTAQVDLQK--SKNEVWADGN 856
QY 731 DTNGKTFNTLKLGE--NGVNITTNRATGTVTFG---IDQSNGLTTPKLTGVGSDTNGNRL 785
Db 857 DSATMTATVRDAKGNLLNDVKVTFNVNSAAAKLSQTEVNSHDIATATLT--SLKNGDYT 914
QY 786 VTEQVPS--ADGNSKTIITKGLSPTLPSIASPSGRNIALGN-----TIEE-----KD 830
Db 915 VTRASVSSGSOANQOVIFIGDQSTAALTLSVPSG--DITVNTAPLHMTATLOOKNGNPLKD 973
QY 831 KSNASIDDLVNLGNFNLKNGKDKQFVSYDVTDFIDGNAT-----TATVTYDE 879
Db 974 KEITFSPNDVASRFSISNSGK-----GMTDSNGTAIASLTGTLAGHTMITARL 1022
QY 880 AN-----QTSKVAYDVN--VDEKTIELT-----GDNGKKOLGV-- 910
Db 1023 ANSNVSDTQPMTEFVADKRAVVVLQTSKAEIICNGVDTELTATVTKDPFDNVVKNLSVVF 1082
QY 911 -----KTIKLTETSGN-----ATFTSTDDDHVAKASDIAGN-- 944
Db 1083 RTSFADPQLSLNARNNTNENGAIEVTLKGVLVGHVHTAEIILNGNRDTKIVNIAPDASNAQ 1142
QY 945 --LNTLAEELHTTKGTANTALOTFTVKKVDENDKADDTNATVKGDKCTSGKVNTLKLCK 1002
Db 1143 VTLNPAQGVV--NNSDSVQLTATVKDPSNHPVAGITVNTMPQDVAAN--FTLE---N 1195
```


Thu Sep 13 14:19:27 2001

Db 291 NVTFAGNLTVDVNRVNRNTPKPIITDNTTRKDNIAISGESAGSMYGRNIKFI----- 345
Qy 362 RKITFKGDDNNSV--ERGLGNTLTIKGDAQTNALTEANIGVVDGNGLKVKLAK-- 416
Db 346 --VTDKGAGVHQGVFAEDDI-NILTDGNSRLNKVYADYVRV-----GKDIELANGQ 398
Qy 417 -----ELTGLTSVATNKIIVS--NTNNAELQSGGLTSPITGK-- 456
Db 399 IHADQOILNATGHVKLNDGSSVSNMNLGISAALNLEATVSNANLSFRVINDTKLNN 458
Qy 457 -----TDKTVYISIDGLKFT-NDNSNIATKGTTRITKKGIFAGTNDGV 498
Db 459 LSKVSARAADLOSGLNLDKA--SVLAHKLTLNLSNDVSLNNOQSKSANKLTKKVRD-- 514
Qy 499 DESKPYLONEKLVKGNSTLN-----SGSLTV--NNTTGNQIOVGANGIK 541
Db 515 ----LNLNNSLSANLTLNNTLNKSKFTAGNLTNVTNVTNLTNDSELAANLTL 570
Qy 542 FATVANNAVANTSATVGTARITEEKICFAGTNDGVDEQAPYLDKERL---KVGRVEITDTS 598
Db 571 L-NVTKNVTLND-----SKLSANKLNDLVNVTNLSKSTLSAGELTPKKVKNYTLNDS 625
Qy 599 GINAGNHIKITGLTNGIANTDAVITIKLDAKP---TLN-----AGDGISINS--NNGD 646
Db 626 ELAANLNLNASHNVTLNKNKLSAOKADIKAVNLTNDTTELTAKNLDINSTITNNGT 685
Qy 647 LVDSSGNIITPTVYNISVKTTLKNSNGTSGNNKFSVSNAH-DNNSLVTAKDLADYLKYNVE 705
Db 686 IAGIFANITTEKLNKEKALIL---AEQNLNFTVNGSHYENKGDIVSKDRATV--TFSK 739
Qy 706 TADSAALPSFKVQNGDN-----SNNAITVGKDTNGKTFNLTALKENGVNITNTRATGV 759
Db 740 NSDFTSGSLVNAQNLKVNYNFTISQD-----ITLIG-----NVTLN-ASGTF 786
Qy 760 TFGIDQSNGLTTPK-LTVGSDTN---GNRLVIEQVPSADGNSTKNIKGLSPTLPSTIAS 814
Db 787 T-----NSGNTLVTKLDVGDIQNFNKGNTLVGEDLHIKSKTITN----- 828
Qy 815 PSGRNIALGNITIEEKSNAASIDVDVNLAFNL-----KNN--GKDKDFVSTYDVA 863
Db 829 -DGKLSIKNL-----NISSEADFNNGLTLGIEALKATKGNFTNKEKAILASNSLL 880
Qy 864 DF--IDGNATTATVYDEA---NOTSKVAYDVNVDEKTIETLDGDNKKOLGVKTIKLT-E 917
Db 881 DISVAEKKTFNNGTIESGKNLNTNTGAF-LNVYDNATI-----RSPGLVNITSTGN 931
Qy 918 TSTNG-----NATFES--TDDHALVKASDIAGNLTNLAEEIHTTKGTANTALQTF 967
Db 932 VSNNGTLISNERLNTISAANFTNESNGTVMSN---GLLNLIAKQGNITNKNLNASROQLN 988
Qy 968 VKKVDENDKADD--TNAITV-----GKD-----GTSKV-----NTLKL 999
Db 989 LTAADNITNDSNISNKTAVLHSLGNLSLNSKDOVYNLGEIYAGNNTSVRAHOLKNDVKL 1048
Qy 1000 KGKNGLDTKT-DKDGTVT-----FGINTOSGLKAGDSTTLNNGISIKNTASNE 1047
Db 1049 MG-----DITTKKEGQASYKLYQASNGHFGNDGSSGSEGD-----LNIG----- 1091
Qy 1048 QIQVGADGVKEFAMVNVVVGAGIDGTTTRITRDEIG--PTGTNGSLDKSKPHLSKDGGINAG 1105
Db 1092 -----KPADLON-----KLTVORIGKIYAGHDLTFNKS-----NAG 1122
Qy 1106 GK-----KIT-----NIQGEIAKNSHDVATGKGIYDLAKTELENKISS 1143
Db 1123 GKSEILNRTINVKNLKLSYSDVSFENNMSQKV-----DLVT--KIFEAKSDIE---L 1171
Qy 1144 TAKTAONSLHEFSADEQGNNTVSNPYSSYDTSKTSVDITPAGE--NGITTKVKNQGV 1200
Db 1172 TFKT--NGHIPVYL-----NFKSNNEKKYRSENKTNFKSIGDLINEALSDSAPEAIE 1223
Qy 1201 RVGIDOTKGLTTP---KLTVNGNNGKGIIVNSQNGTITGLSNTLAN----- 1245
Db 1224 AYYSGSSSNYINPVSYLAALGNAN-----NSSNPHYLNTALKHILGNWQDOLKQENI 1277

Qy 1246 -----VTNDKGSVRTTEOGNLIKDEKTRAASIVDVLISAGFNLOGNCEAVDFVSTY 1296
Db 1278 KVLKOKWEDFKDKGASKMLD---LYPNTDEKAKIAGIIRNGNDTISDVESEDPKKKY 1334
Qy 1297 DIVNFANG---NTTTAKVTYDDTSKTSKVYVDVNVDDTTIEVKDKKLGKVTTLTSTGTG 1353
Db 1335 S--KFQNGEAKNDGTDSYDSTRASEKYKKVENVDHKE-NIDEHKLNIKIGHEITVPGVS 1391
Qy 1354 ANKFALSNQATGDALVKASDIVAHL--NTLSGDITOTAKGASO----- 1393
Db 1392 FEN--LNNKNMHDQPKLGEIDKSIISSELLAQPVYTEASAARSDSPRVNQNDKALDNL 1449
Qy 1394 -----ANNSAG-----YVDADGNKV----- 1408
Db 1450 RFLSYINONNVLGAKYFFENQDLTEDDKLGIKRIGDNYFEHQLITRLIEKVADNHLTLK 1509
Qy 1409 --IYDST-----DNKYQAKN-----DCTVDKTKYKAKDKLVAQAOT-----PD 1445
Db 1510 HGLHDIALVKKLIDSASIOAKDLNLKVGALTKEOKDNLKEDIVWYVYKTEYNGQEVLPQ 1569
Qy 1446 GTLAQMNKSVINKEOVNDANKKOGINE--DN-----AFVKGLEKAAASNDKTKNNA 1494
Db 1570 VYLAKOTIEBEKQRGVGTGQIRAGIIVDKVDDVRNTGTIAGYAVGLE---AKNKLKN-- 1624
Qy 1495 VTVGDL-----NAVAQTPLTFAQDTGTT----- 1517
Db 1625 --TGDILSQRLSKLVGKGLSTGVTVYDGTGATKVKYKARIKSEGHIIYLETDKDNVDLT 1682
Qy 1518 -----AKKLGFT----- 1524
Db 1683 ASELKNGTQIKAKDMLNDIYETSYKYEKLFKNGGELGDRVTQTSQAKSVGTDAF 1742
Qy 1525 ----LTIKG--GQTDNKLTDNNIGVAGTGTGTVKLAKDLTNL-----SVN 1566
Db 1743 DHLHLSLEGDVNQTGSLKANRTTGVVKG--DFNTKAGKDLFHRQIDTIVTSGVYVSASAS 1800
Qy 1567 AGGTIKIDEKGISFVANGQAKANTPVLNSANGLDILGKGVISNVGKGTGKDTDAANVQOLNEV 1626
Db 1801 GGG--QSAGLSLDDQGVETYNKTATAGANAD-----VTNFMKRTRETETSLTHRSEF 1852
Qy 1627 RNLLG-----LGNDNADGNQVN-IADIKDPNPGSSSNRTVTKAGTVLGGKGNNDTEKLAT 1681
Db 1853 NALSCELVMGKADTGGVDINRDVEVITKPEEIAAEQAAEAKKA-EVKENEASETAAK 1911
Qy 1682 GGVQGVKDKGNANGDLNVMVVKTKDGSKKALLATYNAAGQTNVYTNNNPAEADR----- 1737
Db 1912 ETEE--AENDNVAEKDKTKPKFKLTD--BEIAAAPETKGEDEFAAYKAREEDRKKGF 1966
Qy 1738 -INEOGIRFFHVNDQNPVQVQGRNGIDSSAGKSHSVAIGFOAKA-----DGE 1784
Db 1967 TLSAEQIESTKARDEKETTYYELKYGVGAEAENHSAADAISNKARQIIDTQNGLKQDGT 2026
Qy 1785 AAVATGROTAQAGNOSTAIGDNAOATGDQSIATGTCNVVAGKHSAGIDPSTVVKADNSYSV 1844
Db 2027 VAL---QBASDVNLNATGDLAGASAKLKFELST---IEKSRGA-----SDGRSIL 2071
Qy 1845 GNNNOFTDATQDVFVGNNTIVTESNSVALGNSAISANGTHAGTQAKSDGTAGTTTITA 1904
Db 2072 GGRNLNA-ARGGDI--TLNNVETTENSLSLKARDNVNNSGV-TEQKDESNSQSLKVT 2127
Qy 1905 GATCTVKGFAGOTAVGASVGSAGBARRIQNVAAAGVSATSTDAVNGSQL-YKATQ--SI 1961
Db 2128 GASGGGVWAGGCSAG-VSANGVSGS---YNESNTESTSHTNSLLRGKSLRVEAGKDFNL 2182
Qy 1962 ANATNELDHR IHONENKANAGISSAMAMSMQAYIPGRSMVTGGTATHNG---OGAVAV 2018
Db 2183 ISSNVDDVH-LHL-DVKGDTNVVSKQDSYRKRERGVNYSVSAGVGVSTAGGARPNGSVGL 2240
Qy 2019 GLSKLSDNGQWV 2030
Db 2241 GVSAAENNSKIV 2252

[illegible]

```

Query Match          5.1%; Score 526; DB 2; Length 3029;
Best Local Similarity 20.4%; Pred. No. 2.6e-12;
Matches 493; Conservative 269; Mismatches 823; Indels 828; Gaps 111;

QY    60 TLNGSAYAAQQITTTKIEIGQTQNKINNTLKG----- 88
      ||| : | | | | | | | | | | | | | | | |
Db     697 TLDANIATADDIINAIESQAIPITPTGVGGFNFVGDVTTLTVNNDKTFTCVAGAGGLFSINV 756
      ||| : | | | | | | | | | | | | | | | |
nv     89 -----DALATGASGI-----AFGSLSKAAGSQAIAGSVKPDP-----NNGSGNNVGSHA 133

```


QY 1066 VGAGIDGTRITRD-----EIGFTGTNGS-----LDKSKPHLS-----KGINA- 1104
Db 1649 GEMTLDYGTLLADGDEYELSVATNPTGNSATATQIVVDTTAPTVTINAIIVDDIINAV 1708
QY 1105 -GCKKI-----TNISGELAK-----NSHDVATGCKIIDLKTELEN--KISHTAKTAQN 1150
Db 1709 EAGSPVAVSGTTGVEDGQVVVTDIGNTYTATVTCNAWTFNIPVADIANFEAEVVA- 1767
QY 1151 SLHEFSVADEQGN-----NFTVSN--PYSSYDTSKTSVITF--AGE-----NGITTKV 1195
Db 1768 -----TVSDLAGNPATPATRNITVDITAPTVTINAIIVDDIINAVAGSPVAVSGTTTGV 1822
QY 1196 NKG-VYRVGIDQTKGLTTPKLVGNNGKGIIVNSONGONTITGLSN-----TLANVTN 1248
Db 1823 EDGQVVTVTID-----GNTYTATVTCNAWTFNIPVADIANFEAEVVA- 1870
QY 1249 DKGSVRTTEQGNLIKDEKTRAASIVDVLISAGFNLOGNEAVDFVSTYDVFANGNT-- 1306
Db 1871 LAGNPATPATRN1--TVDTVAPAVTIDISDDDTGAQAN-----DFITNDOTLVF-NGTAEA 1923
QY 1307 -TTAKTYIDDT---SKTSKVYDVNVDTTIEVKDKKLGKVTTLTSTGTGANKFALSNO 1362
Db 1924 DSTVVSLDGEIGTGTANCAGEMTLDYCTLLADGDYELSVTATNPTGNSAT----- 1976
QY 1363 ATGDALVKASDIVAHLNLTSGD--IQAKGASQANNSAGYVDADGNKVIYDSTDNKYQA 1420
Db 1977 ATQTIIVDDTTAPTVTINAIIVDDIINAVAGSPVAVSGTTGVEDGQVVTVTIDGNTYTA 2036
QY 1421 KNDGTV-----DKTEV-----AKDKLVAQAQTPDGTLAQMN 1453
Db 2037 TVTGNWTFNIPVADIANFEAEVVA-VDSDLAGNPATPATRNITVDITAPTVTINAIIV 2096
QY 1454 KSVINKEQVN-----DANKKQGINEDNAPVKGLEKA--ASDNK 1489
Db 2097 DDIINAVAGSPVAVSGTTGVEDGQVVTVTIDGNTYTATVTCNAWTFNIPVADIANFEA 2156
QY 1490 TKNAAVTVGDLNVAOPLTFAGDGTCTAKKLGSETITIKGGQPTDK-----LTDNNGV 1544
Db 2157 TEEVATVSDLAGNPATPAT-----RNITVDTVAPAVTIDISDDTGAQANDFITNDTTLV 2212
QY 1545 VAGT--DGFTVKLAKDLTNLNSVNAAGT---KIDEGKISFVDANGO----- 1585
Db 2213 FNGTAEADSTVVVSLDGEIGTGTANCAGEMTLDYCTLLADGDYELSVTATNPTGNSAT 2272
QY 1586 -----AKANTPVLSA-----NGLDLGGKV-ISNVGKGTGD----- 1614
Db 2273 ATQTIIVDDTTAPTVTINAIIVDDIINAVAGSPVAVSGTTGVEDGQVVTVTIDGNTYTA 2332
QY 1615 -----TDAANVOQLNEVRNLLGLGNDNADGNQVNTADIKKDPNCGSSSNRTV 1661
Db 2333 TVTGNWTFNIPVADIANFEAEV-----VATVSDLAGNPATPATRNITV 2378
QY 1662 -----IKAGTVLGGKGNNDTEKLTATGGVQVG-----VDKDG- 1693
Db 2379 DTTAPTVTINAIIVDDIINAVAGSPVAVSG-----TTTGVEDGQVVTVTIDGNTYTA 2431
QY 1694 -----ANGDLSNVWVKTKQDGSKALLATYNAGQTNVYTN-PAEA 1734
Db 2432 TVTGNWTFNIPVADIANFEAEVVA-VDSDLAGNPATPATRNITVDTVAPAVT 2485
QY 1735 IDRINQO--GIRFFHVND-----GNQE--PVQGRNGID-----SSAGSKHSVAIGF 1777
Db 2486 IDISDDTGAQANDFITNDTTLVFNTEADSTVVVSLDGEIGTGTANCAGEMTLDYTG 2545
QY 1778 QAKADGEAAVAIGRQTAGN-----QSIAIGDNAQATGQDSIAI-----GTGNVA 1823
Db 2546 TLLADGDYELSVTATNPTGNSATATQIVVDTTAPTVTINAIIVDDIINAVAGSPVAVS 2605
QY 1824 GKHSGA-IGDPSTVKAD-NSYS-----VGN-----NNOFTD--ATOTDVF- 1860
Db 2606 GTTGTGVEDGQVVTVTIDGNTYTATVTCNAWTFNIPVADIANFEAEVVA-VDSDLAGNPA 2665

QY 1861 --VGNNTVTES-----NSVALGNSAISAGTHAGTOAKKSDGAGTATTTT 1903
Db 2666 TPATRNITVDITAPTVTINAIIVDDIINAVAGSPVAVS-GTTTGE-----DQGVVTVTI 2720
QY 1904 AG-ATGTVKGFAGQAVGAVSVGASGAERRIONVAAGEVSA-STDVANGSSQLYKATOSI 1961
Db 2721 DGYTYTATVTCNAWTFNIPVADIA-----NFEATEEVVATVSD-LAGNPATPATRNI 2771
QY 1962 -----ANATNELD 1969
Db 2772 TVDTVAPAVNELD 2784
RESULT 12
T31102
filamentous hemagglutinin 1 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31102
R:Ward, C.K.; Lumbiey, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326
A:Accession: T31102
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4152 <WAR>
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1
C:Genetics:
A:Gene: lspA1
Query Match 5.0%; Score 513; DB 2; Length 4152;
Best Local Similarity 20.4%; Pred. No. 1.2e-11;
Matches 483; Conservative 344; Mismatches 953; Indels 590; Gaps 113;
QY 2 NHIYKVFIPNKATGTFMAVAYAKSHSTGGSCATGQGVSVRTLSFARIAALAVLVIGATL 61
Db 3 NKRYKLFISKVKNCPLVPAENIKSAGNSGSKIAEDOEEDPSLA----- 51
QY 62 NGSAYAQIITKTEIGOTNKN-NLTKGDALATGEASIAFGSLSKAGGSOAIAI-GSVKP 119
Db 52 ---CSLSPSSSTHGLHNSPLKVPKGLSVLLSLMPATPLLAAQNTAEALNGKVV 108
QY 120 DPNNGSNGVNGSHAKNESTAGDGLAEGDASITAGSDDLYLPLKNL-----DKNEF 172
Db 109 DSQHSSTRIY--EQKNDNSKDGIVVVEIANPEVDGVDNRKEFNIPNSAVENSRTE 166
QY 173 HKLIHGHEILUKKIQTSDGKIKYRTRAQCHASTAVGAMSYAOGHFSNAGTGTATAAAY 232
Db 167 TSQLVG-KLHANIQLOKEAKLIINOVTDG-HESNIOGALEVA-----GKKADLIIVN 216
QY 233 SLAVGLAAQATKOSSIAVGSNA-----KANAFATAIGGNTVNLGRGVALGFGSQILDR 287
Db 217 PNCITLNGVKTINTDFPVVSTSDLIIPHRENGLLSVR-NGKVTIDKG-GVATNGLSHEVV 274
QY 288 DNNITASAYVPLGKTLADQYKATRGDSTDFISIGNNNNNSSIRRKIIINVGAG-SRDT 346
Db 275 ARNIDQKGTIVAKT--ENQKSYNPANIT--FAAGSLNLYNKLTREATPPISSGTSRTSDTP 330
QY 347 AVNV-----AQLKLYBELANRKTFFKGDGDNNSNSVERGLNTLTIKGDAQTNALTE 398
Db 331 AISADSAGSMYGNIKFV-----VTDKAGVKKHGII-----FSE 365
QY 399 ANIGVVTGNGLVK-----LAK--ELTGLTSVSATNKTIVSNT----- 435
Db 366 NDINIKMDGNGASLKELYAKKDIDILAKOIELEKGOLOQANNKILLNSTKINRINASEV 425
QY 436 --NNNNAELSGGLTSPITGTTKTDKTVYSDGLKFTNDSNSTATGTTTRITKKIGFAG 493
Db 426 SADNVVVKSENALENASMSANSLDIVITVKIE-----VNRSSKVSVA-GTANIKASNI--- 476
QY 494 TNDGVDSEKPYLDNEKLVKGNSTLNSGSLVNVNTTGNKIQOVGANGLIKFPATVANNVANTS 553

Thu Sep 13 14:19:27 2001

Db 477 -----TLOGSSVANKIYLN-----VTNATLNQSKLSAKDMEL-----NVTHNI 517
QY 554 ATVGTPARITEKIGFAGTNDGVDQAPYLDKERIKYGRVEIITDSGINAG-NHKITGLTN 612
Db 518 TLNNTSKLSAQKANIENLTNGEAS-LVAEKLDINAIDKITNNGTIAGLTANIT--TK 574
QY 613 GIANTDAVITKOLDAKPTLNA-----GGGISIN-----SNGDLVDSGNGITPTTN 660
Db 575 ALENDRALILAHONLFTVNGSHYVKNKGLIVSKDAIVTFESNNSDFTSGSKLVDAQNN 634
QY 661 ISVKTTLKNSGTGNNKFSVNAHNNLSVTAKDLADYLKVNVEADTADSLPFEKVQNGD 720
Db 635 LIVNVN--NFNITOGSEII-----LHGN---VTLNAGNFTNSGMLTMMKELNITSNFI 685
QY 721 NSNNAITVKTNGTFTNLTALKGNGVNIITNRTATGTVTFGIDQSNGLTTPKLTGVSdT 780
Db 686 NAGN-LTTGKNLEVHSNTTVKNDGKL-----LHGN---VTLNAGNFTNSGMLTMMKELNITSNFI 722
QY 781 ----NGNRLVIEQVPSADGNSTKNIKGLSPLPSIASPS-----GRNIALGNITIEKDK 831
Db 723 DFTNNGTLGLLEALKTASGNFTNASNG-----SLASNKSLDIYGNFTNNGTIESVKS 776
QY 832 SNAAS-----IDVNLNAGFNKNGKDKDFVSTYDVFIDGNATTTATVYDEANOTSKVA 887
Db 777 LNTNNTYFINNATIKSYGVLAITSOGFTN-----DSNGT--VMSHDLNITSQ-- 824
QY 888 YDYNDEKTIETLTDGNGKQKLGKVIKTIKLTETSTNGNATTFSTDDHALVKA-SOI----- 941
Db 825 --ANIINKNL-LAGGOG-----LNLJT---AKGNITNDSNSTAIVLHNSNNLNAN 870
QY 942 -----AGNLTIAELH-----TTKCTANTALOTFTVKKVDENDKADD 979
Db 871 NKVYNGEISYQAGNISVEAKLHNDVKLSGNTTTTKSGNATVKNISI-----GGGLHD 925
QY 980 TNAITVCKDGTCKV-----NTLK-----NLGDIKTDKG-----TVTFG 1018
Db 926 ANSLRVGELLTLNGKFAIDLNLQKVALRGKIYAGSNLTFRKA-KEGEKEQSTAAQKIING 984
QY 1019 -INTQSLKAGDSTTLNNGSLSTKNTASNEQIOVGADGVKFAVMNVGVGAGIDGTTRIT 1077
Db 985 TINVKNKLEYSNVVDENNRSMQ-VNLXEKIFNGDNPTLTL-KNGVTFK-DFSNRRR 1041
QY 1078 RDEIGFTGTLGSLDKSPHLSKDCGINAGGKKITNIOSGETAKNSHDAVGGKI---YDL- 1133
Db 1042 RASNDGEGTINKTDFNVHLLIEAF-----SGSYNGNDHRASTDGHVKSPYLL 1090
QY 1134 -----KTELENKISSTAKTAQNSLH--BFSVADEOGNFTVSNPYSSV----- 1174
Db 1091 VLAQAVNNTGEGNYL---KTALQHIFGPNWMDLTNTTNDTITINDKNQKLKWEKFN 1146
QY 1175 -DTSKTSVITFAGNGI-TTKVKNKGVVRVIGDQTKGLTTPKLTGVNNNGK----- 1223
Db 1147 GENNHSINLIYPADEGEVAKIIFAGVLNNG---TNGVEDKVVYQELNDKAKKEYEDKFAK 1203
QY 1224 ----GIVINSQNGQNTITGLSNTLIANTNDK-GSVRTTEQGNIIKDE----- 1265
Db 1204 KEOGRFKSRFQNGEFDWAG---DWAKEGNEISYGSKEETEETKNGIKKEHTVNIKGHEIKVP 1260
QY 1266 -----DKTRAASIV-----DVLISAGFNLOGNEAVDFVSTY 1296
Db 1261 TVSFENLNINHOQKSDGIDKSIISSELLAQPIYVAKADVPDVPDPRVAQNDAKVEDGLY 1320
QY 1297 DT-VNFAN-GNTTAKV-----TYDDTSKTSKVVYDVNVDDTTI-----EVKDKLGVK 1343
Db 1321 RTRLSYVNNQNYLGAKYFFNQDTEDDKLGKIKRIGDNYFEHOLITRLEKVAADNHLTLK 1380
QY 1344 TTTLTSTGTGANKFAL-----SNOATGDAIVK-----ASDIVAHLNT--- 1380
Db 1381 -----HGLHDIALVKLLIDSASTQADLNLKVGELTKKEDKONLKEDIVVYKTEVN 1432
QY 1381 -----LSGDIOTAKGASQANSAGYVDADGNKVIYDSTDNKYYQAKNDGTV 1426

RESULT 13

H81193
hemagglutinin/hemolysin-related protein NMB0493 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81193
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.B.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Reference number: A81000; MUID:20175755
A:Accession: H81193
A:Molecule type: DNA
A:Residues: 1-2703 <TET>
A:Cross-references: GB:AE002405; GB:AE002098; NID:g7225708; PIDN:AAFA0927.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0493

Query Match 4.9%; Score 512.5; DB 2; Length 2703;
Best Local Similarity 20.0%; Pred. No. 7.2e-12;
Matches 531; Conservative 345; Mismatches 893; Indels 887; Gaps 134;

QY 1 MNH-IYKVIENKATGFMVAEYAKSHSGGSCATGQVG-----SVRTLS 45
DB 1 MNTLTVKVFNHRNCRMIATAAARE--GKNTADTQAVGILPNDIAGPAGFIHSIVS 58
QY 46 FARI-----AALAVLIGATLNG-----SAYAOQITTKIEIG-----QTNKINNTLKGDALA 92
DB 59 FSLISLLGSALILTSSSAQAQIVADKAPAOQOPILOTGNGIPQVNIQTPTSAG---- 114
QY 93 TGEASTAFGLSKAGCSQAIAITGSVKPDNNGSNGVNGSHKAGNESIATGGV-----IA 147
DB 115 -----VSVNOYAO-----DVGNRGAILNNSRSNTQTQLGWIQGNPWLA 154
QY 148 EGDASTAGSDDLPLPKNLDAKNEPHKLHGHELLKTIOTSDGKIKYRRTAOGHASTA 207
DB 155 RGEARVV-----NOINSSSSQLNGYIEVGGRAEVVIAMP 191
QY 208 VGAMSYAOGHFNAFGYATATAEAYSLAVGLAAQATKQSSIAV---GSNAKANAFATAI 264
DB 192 AGIAYNGGGFINASRATLTTAQPOYO-AGDLSGFKIRQGVVVIAGHGLDARDTDY----- 245
QY 265 GGVTVVNLGRVALGFGSQILDRDNDNTDASAYVPLGKTLADQYKATROGSDTDF----- 319
DB 246 -----TRILSYHSKIDAPVWGQDVVVAGQVNDVAATGDAHSPILNNA 288
QY 320 --SIGNSNNSIRRKIINVAGSRDTPAVNVQAOLKVEELANRKITEFGDGDNNNSV 377
DB 289 ANTSNNTANGTHIPLFADTG-----KLGGMYAN-KITL-----ISTV 326
QY 378 ERGLGNTLRIKGDATNALTEANIGVVTGNGLKVKLAKELTGLTSVSNATNITVSNTNN 437
DB 327 EQA-----GIRWQGWFA-SAGNVAVNAEG-----KLVN-TGM-IAATGE-----N 364
QY 438 NNAEQSGGLTSPITGTDKTVVYSIDGLKFTNDSNSTATGTTTRITKKKIG-FAGTND 496
DB 365 HAVSLHARVNHNSGTVASODDAIHS-----QTLDSNGTVLSSG--RLTVNLGRKLQNQN 418
QY 497 G-----VDESKPYLDNEKLKVCNSPL-----NSGSLTVNNTTGNKQIQV 535
DB 419 GTIOAARLDMSTGGLDN-----TGNITQTSQALDLVSAGKFDNSGKIGVSDVP-----QT 469
QY 536 GANGIKFATVANNVANTSATGATRIEKGIFAGTNGDVGEOAPYLDKRLKVGREIT 595
DB 470 GLN--PNPSVPIQIPSTATGSGSSTVSVK---PGSNPVSPTAP---AKNAVAGRIQTT 521
QY 596 ----TDSGINAGNHKITGLTNGTANTDVTIKOLK-----DAKPTLNAGDGISINS- 642
DB 522 GAFDNAGSINAGQIDIAAQNGLNGSLNAAKLRVSGDSFNFNTVKGLQAHD-LAVNTQ 580
QY 643 ----NNGDLVDSSCNTITTPYN-----ISVKTWKLNSNGTSGNNKFSVNAH----- 685
DB 581 TAKNSGHLTQTKIDNRELHNAGEATAANLTIHSGRLSDNKKGNIRAAHLQDTPAGLH 640
QY 686 -----DNNSLTAKDLAD-----YLNKVNETAADSA-----LPSFKVQNGDNNNN 724
DB 641 NAGNILDASCTVTKNLRNTGKVSVARLNTBQOTLDNTRGRIEATVNIQSOLTNQSG 700
QY 725 AITVGDKTNGKFTNTLKLKNGENVNTTNRATGTVTFGIDQSNQ---LTPPKLTIVGSDTNG 782
DB 701 HITA-----TEQLTINSRVNDQNGKLLSANOQAQAVSDGLYNQHEIATNRQLSI-HDKNQ 756
QY 783 NRLVIEQVPSADQ--NSTKNIILKGLSPTLSPASPSGRNIALCNTIEBKDKSNAASIDQ- 839
DB 757 NTLALN-----NADGTIOSAGNV-----SLQAKSLANNGLTLAGNKLDIALTDFF 801
QY 840 -----VLNAGFNLNKNGKDKDFVSTYDVFIDGNATTATV 875
DB 802 VVERDLTAGQLNLSTIKRLKNTHTLQAGHTLKLNAGN-----IDNQVTGKII 849

QY 876 TYDEANQTSKVAYDVNVNDEKTIETLGDNGKKQLGVKVTIKLTETSTN---GNATTFST--- 929
DB 850 GGEOTDITS-----EQIVDNR---GLNSDGLTHIGAGQ-TLTGTGTGKIYHIALDAQIL 902
QY 930 ----DDHALVKASDITAG--NLNTLAEETHHTKTGTANTALQTFV-KKYDENDKADDTNA 982
DB 903 LAREETTESGTAGATAARKRLDIGAKEIHNQEGALLSSEGIFAVGNRLDEQIHA----- 957
QY 983 ITVGKDGTSKGVNTKLKNGKGLDIKTD-----TVTFGINTQSGLKA----- 1010
DB 958 -----AGMADTF-VNGSAGLEVQGDALMSVRRNQNNINNHKFTETYLAKAEKQVRDIT 1008
QY 1011 -----KDG-----TVTFGINTQSGLKA----- 1027
DB 1009 VLQNTYYOAGKDGFLDNSQGQKQDQTTATFHLKNGSRIEANQHWYRDYHITYKERIEN 1068
QY 1028 -----GDSITLNNNGLSIKNTASNEOIOGA-----DGVKFAVNVGVVAG- 1069
DB 1069 RPAHITVGGDLTASGQNWLN-----KDSRIVVGGRIITDOLNKEITNOSTTGKGRDVA 1123
QY 1070 -----IDGTRITRDE-----IGFTGTNGSLDKSKPHLS 1098
DB 1124 GTQWDSVTKKGWYSGRKQRORRTERNHPTPHDTQTLFTHDFTPVSVIOQNAASFPQPAAS 1183
QY 1099 K-----DGINAGKKITNIOSEIATKNSHDAVT--GKIYDLKTELENK----- 1140
DB 1184 AIKLDIGVSTAANGORIHTGVNVSLNNATVLPNSSLY--TTHPDNKGWLVDTPQFAD 1241
QY 1141 ----ISSTAKTAQ-----NSLH-----EFSVADEQGNFTVSNPYSYDTSKTSDDVI 1183
DB 1242 YRWLGSYMLQOLOLDTNHLHKLRGDGYEQKLVNEQIHQUTGYRRLDGY-----RSDEE 1297
QY 1184 TFAG--ENGITTKVNGKV--VRVGIDQTKGLT-----PKLT 1216
DB 1298 QFKALMDNGLTAAKTGLTPGIALSAEQVARTLSDIVVMENQTVLSDGSTQTVLPVKY 1357
QY 1217 V-----GNNNGK-----IVINSQNG-----QNTITGLSNTL--ANVTNDKGSVRTTEOG 1259
DB 1358 ALARKGDLNTSGGLISAEQVLLKLQNLNLTNSGTIAGRAVLQIARNINSNGIQADQIG 1417
QY 1260 ----NIKDEKTRAASIVDVLVSAGFNQNGEAVDFVSTYDVFANFANGTITAKV--- 1311
DB 1418 LKAESINIDGGVQAGRLTAQAQINLNG-----TTQTSGERNENGTADRMAGI 1469
QY 1312 ----TYDDTSKTSKVY-----DVNVDDDTI--EVKDKKLGVKTTTLTSTGTGANKF 1357
DB 1470 NVVGSHTEQVDNRTSGILSLHASDINLNAATVSNQVQD-----GTTQITAGNNLNG 1523
QY 1358 AL---SNQATGDALVKASDIVAHLNLTSGDIQTPAKGA-SOANNS-----AGYVDADGNKVI 1409
DB 1524 TIRTEHREAYGTLDDEHRRHVRQSTEVGSSIRTPONGALLRAGNDLKIROGELEAEEGKV 1583
QY 1410 Y----DSTDNKYQ-----AKNDGTVDKTK-----EVAKDKLVAQA 1441
DB 1584 LAAGROVTISEGRQITELDTSVSGSKGILSSTKTTHDRYRFSHDEAVSGNIGGKMLVAA 1643
QY 1442 Q-----TPDGTLAQ-----MNVKSVINKEQVND--ANKQGINEDN--AFVKGLE 1482
DB 1644 GQDINVRGSLNISDKGIVLKAGHDIDISTAHNRYTNEYHESKSGVMGTGGLFTIGNR 1703
QY 1483 KAASDNKTNAAVT---VGDNLNAVAQTPLTTFAGDTGTTA-----KKIGETLTIKGGQDFTN 1535
DB 1704 KTTDDTDRNTNIHVTGSIIGSLN-----GDTVTVAGNRYRQGTSTVSSPEGR---N 1750
QY 1536 KLTDDNIGVYVAGTGDGFTVKLAKDLTLNLSVYNAGGTKIDEKGISFVDANGQAKANTPVLSA 1595
DB 1751 TVTAKSIDVEFANN-----RYATDYAHTQ-----EOKGLT-----VALNVPPVQA 1790
QY 1596 --NGLDLGKVISNVGKTKDRTDAAN-----VOQLNEV--RNLIGLCNDNADGNQVN 1643
DB 1791 AQNFIOAQNVGSKNKRNVNAMAANAANWQSYOATQOMQOQFAPSSSAGOGONNQSPSIS 1850
QY 1644 IADIKKDPNSGSSSNRTVIK-ACTVILGGKGNNDTEKATG-GVQ-----VGVDKGNANG 1696

A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ydbA_2
A:Start codon: GTG

Query Match 4.8%; Score 499.5; DB 2; Length 2020;
Best Local Similarity 20.8%; Pred. No. 1.5e-11;
Matches 405; Conservative 224; Mismatches 663; Indels 657; Gaps 100;

Qy 284 ILDRNNNTDASAYPLGKTLADQ-YKATROGSDTDFISIGNSNNNSSIRKLIINVGAGS 342
Db 147 ILDKTKET-----LITRDSVFTYENADGT--ISLQDSNGRKATL----- 184
Qy 343 RDTDAVNVAAQLKVLAEELANRKITFKGDGDNNSNSVERGLNLTIKGDAQTNALTEANIG 402
Db 185 -----NLWQI-----DEANNTVALEGSADGATKQYNHNGELVITGD---NATVNNNGK 231
Qy 403 VYTDGNGLKVKLAKELTGLTSVSATNKITVSTNNNAELOSGLTFSPITCTTKTDKTVY 462
Db 232 TTVDG-----KDSGTG-TEIN-----GNGKVIQDGLDVS----- 261
Qy 463 SIDGLKFTNDSNSIATKGTTRIT-KKIGFAGTNDGVDESRYLDNEKLKVGNSLTNSGS 521
Db 262 GGHGIDITGDSATVDNKGTMVTVDPESMGIQ-----IDGDKAIVNE-----GESTITNGG 312
Qy 522 LTVNNTGNKQOVGANGIKFATVANNVANTSATGTARITEEKIGFAGTNDGVDEQAPY 581
Db 313 -----TGTO-----INGDD--ATANNGKT-----TVDGKDSGTGE 341
Qy 582 LDKERLKVGRVETITDSGINAGNH--KITGLTNGTANTDAVTIKQLKDAKPTLNAGD--- 636
Db 342 INGN--GKVIQDGLDVSNGGHDITG-----DSATV-----DNKGTMTVDPES 386
Qy 637 -GTSIN-----SNNGDLVDSSNITPTPINISVKTKLN-SNGTSGNNKFSVSAHDNN 688
Db 387 IGIOVDGDQAVVWNEGESAITNGG-----TGTQINGDDATANNNGKTVTDGKDS- 435
Qy 689 SLVTAKDLADYLNKVNETADSLPSFKVQNGDNNNAITVGKDTNGKTFNTLKLKGENGV 748
Db 436 ---TGTEIAGNGKV-----IQDGD-----LDVSG-----GGHGI 462
Qy 749 NITNRRAT---GTVTFGIDQSNCL-----TTPKLTVGSPTNGNRLVIEQ 789
Db 463 DITGDSATVDNKGTMVTVDPESGIQIDGQAIWNNEGESTITNGGTGTQINGN----- 516
Qy 790 VPSADGSKTKNIKGLSPTLPSIASPSG-----RNIALGNTIEEKDK- 831
Db 517 -DATANNSGKTTVDGKDSGTGKTAGNIGIVNLGSLVTVGGAHGVENIGONGTVNNKGD 575
Qy 832 --SNAASIDDLVNA-GFNLKNGKDKDFVSTYDVFIDGNATATATVYDEANQTSKVAY 888
Db 576 VVSDTSGISGLVINGEGATVSTGDVN-----VSNEATGFSITTSNGKVSLAGSM 624
Qy 889 DVNVDEKTIELTGDKNGKKOLGVKTKLTETSTNGNATFTSTDDHVALVKASDIAGNLT 948
Db 625 QVGFSTGVDLNGNNNSVTLAAKDLKVVGQKATG-----INVSGDANTVNT 671
Qy 949 ABEIHTTKGTANTALQTTVKKVDENKADDTNATVTK-----DGTSGKVTTLKLG 1001
Db 672 GNVLDKDKTADNAAEYFDPVSGINNYGSDNNVTLDGKLTVSDSEVTSQSOLF----- 727
Qy 1002 KNGLDIKTDKGTVTFFGINTOSGLKA-GDSTTTLNNG-----LSIKNT-ASNEQIOVGADG 1055
Db 728 -----DGSA-----EKTSGLVVIGDGTVNNNGGLELIGEKALADGQVTSRTG 773
Qy 1056 VKF--AMVNVGVGAGIDGTTTRITRD-EIGFTGTNGSLDKSKPHLSKDGINAGGKKITNI 1112
Db 774 YSVTSVIVVSGESSVYLVNGDTTISGEFFPLGFAGV-----IRVQDKALLEI 818
Qy 1113 QSG-----EIAKNSHD-----AVTGGKIYDLKT-ELENKISSATAQNSLHEFS 1156
Db 819 GSGATLTMDIDISFEHHGTRTPELTIVADSGAKIVNKGTVFION----- 861

Db 1851 VSIYTGKSRNEQKRHYTEAASQIIGK--QTTLAATGSGEQSNINITGSDVIGHAGT 1908
Qy 1697 DL-----SNVVKTK-DGSK--KALLATYNAAGTQNVVWNNPAAEIDRINEOGIRFFHVND 1750
Db 1909 ALIADNHIRLQSAKQDGEQSKNKSQWN-AGVAVKIGN-----GIRFGITAG 1955
Qy 1751 GN-----QEPVVGRCIDSSASGKHSVAIFQAKADGPAAVAIQROTOAGN-----Q 1798
Db 1956 GNICKGKEQGSSTHRTHTVSGTGTGKTIIRSGDGTTLKGVOLIGKGTQADTRNLHIESVQ 2015
Qy 1799 SIATGDNAQATGDSIAIGTGNVAGKHSAGIDPSTIVKADNSYSVGN--QF 1850
Db 2016 DTETYSQKQNGCNQVTVVGYFSAGSYR-----QSKVKADHASVTCQSIYAGEDGYOI 2070
Qy 1851 TDATOTDFG-----VGNNI-----TVT-----ESNSVALGSSAISAGT 1885
Db 2071 KYRDNITDLKGLIYSSQSAEDKGNLFOTATLTASDIQNHRSRYECRSFGIGGSFDLNGW 2130
Qy 1886 HAGTOAKK-----SDGTAGTITTTAGATGT-----VKGFAGOTAVGAVSVGA 1926
Db 2131 D-GTVDKQGRPTDRISPAAGYSGDGSKNSTTRSGVNTNHIHITDEAGQLA--RTGRTA 2187
Qy 1927 SCAERRI-----QNVAAAGEVS-----ATSTDA 1948
Db 2188 KTEARIYTGIDETADQSHGLKHSFDPKDAVAKEINLQREVTKFGRNAAQAAVAADK 2247
Qy 1949 VNGSLYKATQSIANATNELDRHIQNEKNKANAGISSAMAMSPQ--AYIP----- 1998
Db 2248 LGNTQSYERY--EARTLLEAELONTDSEA-----EKAAFRASLGQVNAVLAENOSRYDT 2300
Qy 1999 -----GRSV--TGGTATHNGOGAVAVGLSKLSDN-----GQWVEKINGS 2037
Db 2301 WKEGGTGRSLTHGAAGGLTGTSLGGLAGGTSLAAPYLDKAAENLGPAGRAAVNALGGA 2360
Qy 2038 -----DTQGHVGAAGA 2049
Db 2361 AIGYATGSGGAVVGA 2376

RESULT 14
C48399
ABC-type transport protein ydbA_2 - Escherichia coli
A:Species: Escherichia coli
C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999
C:Accession: C48399; D64891; H64891
R:Moszer, I.; Glaser, P.; Danchin, A.
Biochimie 73, 1361-1374, 1991
A:Title: Multiple IS insertion sequences near the replication terminus in Escherichia coli
A:Reference number: A48399; MUID:92190338
A:Accession: C48399
A:Molecule type: DNA
A:Residues: 464-2020 <MOS>
A:Cross-references: GB:D85081; NID:g3041754
A:Experimental source: strain K-12
A:Note: sequence inconsistent with the nucleotide translation
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-839, 'LDLPLYFQTSVIT' <BLA1>
A:Cross-references: GB:A600237; GB:U00096; NID:g1787665; PIDN:AAC74483.1; PID:g1787668;
A:Experimental source: strain K-12, substrain MG1655
A:Accession: H64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'M', 915-2020 <BLA2>
A:Cross-references: GB:A600237; GB:U00096; NID:g1787665; PIDN:AAC74487.1; PID:g1787672;

QY 1157 VADEGNNFTVSNPYSSYDTSKTSVITPA-----GNGIITTKVKNKGVVRVIGIDQTKGLTTP 1213
Db 862 -----LGFATVTGEN--TTGINSGISLQKNGKDPAPSP 893
QY 1214 KLTVCNNNGK-----GIVINSQNGQNTI-----TGLSNTLANVTNDKGSVR--TTEOGNIK 1263
Db 894 IVLLATNGGSATNAGTIYCKVTEQHSVFNKYSTGINSFI--FNNDVSSITGLVLAOSN--- 949
QY 1264 DEDKTRAASIVDLVSAGFNLOGNEAVDFVSTYDVFNFANGNTTTAKVTYD--DTSKTSK 1321
Db 950 -----STIINTDSGIIIDLYCRG--SVGMLATADSTAENQKGTILDSMWVDANDITAMRD 1001
QY 1322 VYDVNVDDTTLEVKDKKLVKVTTLTSTGTGANKFALSNOQTGDAVKASDIVAHLNLT 1381
Db 1002 IASNSAIDFGT-----GVGVGTSYSGAGKNATAI--NQLGG-----VITIYNAG 1044
QY 1382 SGDIQTAKGSAQNNAGSYVDADGKNVIYDSTDNKYQAKNDGTVDKTKVAKDKLVAQA 1441
Db 1045 AG--HAAYGAS-----NVTINQGTIN-----LEKNGYDSS--LAANTLVGMA 1083
QY 1442 QTPDGTLAQNMVKNVINEQVNDANKKOGINEDNAFVKGLEKAASDNKTNAAVTVGDL- 1500
Db 1084 VYEHGT-----AINDQTGVINVTGT-----QAFYNDGTGTIVNYGTIC 1123
QY 1501 -----NAVAQPLTFAGDTGTAKKLGTELTIKGGQDTYKLTDN--NIGVVA 1546
Db 1124 TFGVCQSGNEYNNTDFTSIYTG--GDTITRSGETVTLNKAASVTDKLAGVNVNSGTLS 1181
QY 1547 GTDGTFTVLAKDLTNLNSVWAGS---TKIDEKGISFVDANGQAKANTPVLASCLDLG 1601
Db 1182 G-DQITVS-----SGLENTSGGIINLVKLDK-----GAVIKNAGVMT--NNVDVS 1225
QY 1602 KRVISNVGK-----GTDKTDAAVQQLNEVRNLLGLNDADNGOVNIADIKKDPN 1652
Db 1226 GGILNAGEMTAQITWAGADSSLVNNTCTINKIQNAGVEN-----N 1268
QY 1653 SSSSNRTVIKAGTVLGGKGNNDTEKLTGTVGVGVVDKGNANGDLSNVW-VKTOKDGSK 1711
Db 1269 SGSVTGRMM-----SAGGVFNQTD-----GAIMRGAALTGTAVANNEGTPWNLGSSEGN 1319
QY 1712 KALLATYNAAGQTN---VYTNPAIDAIDRINEQGTREFHVNDGNQRPVVGVRNGRIDSSAS 1768
Db 1320 TCMLEVNNSAFNNRGEFILNDKNAV--HINOSGTLY---NTGHM-----NISN 1364
QY 1769 GKHSVAIGFQAKADGEAAVAIGROTQAGNQSTAGDQAQTDQSTAGTGNVWAGKHSG 1828
Db 1365 SSHGAVNMW-----GNGRFEINDGTIDVSAKSLV----- 1394
QY 1829 ALGDFSTVKADNSYSVGNNNQTFDQTDVFGVGN---ITVTSNSVALGNSAISAG 1884
Db 1395 -----VSANNA--GQNAFFWQDNGVINFDHDSASAVKTHSNFTA--QNDGIMNI 1442
QY 1885 THAGTOAKSD-----GTAGTFTT-----ACAT-----GTVKGFAG 1915
Db 1443 STGAVAMEGDKNAQLVNNGTINLGTAGTFTDGMQMLDANATADAVIENNGTINIFAN 1502
QY 1916 QTAVGAVSYGASGAEERRIONVAAGEYSATSTDAVNGSQIYKATQSI-----ANATNELDHR 1971
Db 1503 DSFAFSV-LGTVGH-----VYNNCTVVIADGVTCGLIKOGDSINVEGMNNGNCSSE 1554
QY 1972 IHQNEKANAGISSAMAMASMPQAIIPGRSMYTGII-----ATHNGOGAVAVGLSKLSDNG 2027
Db 1555 VHYGD-----YTLDPVPKPTVSVTSGSDEAGGSMNINLVVVGNT----- 1595
QY 2028 QNVFKINGS-----DTQGHVCAAVGAGF 2051
Db 1596 -----VNGSAGKLVNNASMMNGVEINTGF 1619

RESULT 15

CB3339

hypothetical protein PA2462 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: CB3339
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Accession number: AB2950; MUID:20437337
A:Accession: CB3339
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5627 <STO>
A:Cross-references: GB:AE004673; GB:AE004091; NID:g9948501; PIDN:AAG05850.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2462

Query Match 4.7%; Score 487; DB 2; Length 5627;
Best Local Similarity 20.8%; Pred. No. 1.8e-10;
Matches 488; Conservative 319; Mismatches 875; Indels 668; Gaps 111;

QY 84 NTLKGDALATGE--ASI-----AFGSLSKAQSQAIAIGSVKPDPPNNGSNGVNGSHAKGN 136
Db 1725 NSVRGRIAAAGDLHASVTAFNQAGGELS--SEGALMLEADSL-----DNRSGGLVS--ADGN 1777
QY 137 ESTAI-----GGDVLAEGDASI-----AIGSDLYL--PKNLD----- 167
Db 1778 LVVSARRIDNRAGETASPCQVTLIDVAEQLDNRGGKAGISGLRLAAPRVLNQDGGVLA SR 1837
QY 168 --LKNEFHKLHGHHEILKKIQTSTD-----GKIYRRTRAQGHASTAVGAMSYAQGHF 218
Db 1838 DGLRLNGAELEFNGLLSQSQSIDVLGVLCN-----QAGSLSQGRLSVKSGRL 1889
QY 219 SNAGTYATAEAAYSILAVGLAAQATKQSTAVGNSAKANAFATATGNTVNLGRGVAL 278
Db 1890 DNOGGAVSSA-----GTLSSQAGALNNOGGRVVTDAAGAVLR- 1926
QY 279 GFGSQILDRDNTDASAYVPLGKTLADQYKATRGQSDTDFISGNSNNSSIRRKIINV 338
Db 1927 ---SASLDSNGQIYSA-----KGAEE-----IRGSLNNSQK-----GI 1959
QY 339 GAGSRDTDVAVNAQLKVLBEA-----NRKITFGKGDGNNNSV-ERGLGN-----TLT 386
Db 1960 GSG-----AGLAVADLVDSNQRITAKGALDANLKLQDQSGRLVSDTAIALD 2010
QY 387 IKGDAQTN-----ALTEANIGVVTGNCILKVLAKELTGLTSVATNK----- 429
Db 2011 LRGGELVNRQAQGLIATPGALLRLQLGVVDNSGGGEISSDRSET-LAATALSNRGGRVISG 2069
QY 430 ----ITVSVNTNNNAE---LQSGGLTFSPITGKTDKTVISIDGLFTDNDNSIATKGT 482
Db 2070 DSILTRIAALDNLQGVLSASGGL-----DVAALVFDNHSIGVASKGDT 2114
QY 483 RITKKKIGFAGTNDGVDESKPYLDNEKLKVGNTSLNGSLTNTVNTGNKQI-----OVGA 537
Db 2115 HI-----GVNR-----LENEAGRV-----VSEGALDLT-----AKQVSSAKGRIA 2150
QY 538 NGIKFATVA--NNVANTSATVGTARITEKI-----GFACNTDGVDEQAPYLDK-----ERL 587
Db 2151 KGDQVQVTLTBOQGGELASQGTTLTLDADSLDNRNGGLVSADGGVTAERQIDNRGETS 2210
QY 588 KYGRVEITTDGGINAGNHKITGLTINGIANTDAVTIKO-LKDAKPTLNAGDGISINNGD 646
Db 2211 SVAKVALAVREQLDNRGGKVG-----DSELSLTVORLLNQAKGVLSARDGLHL--DGAE 2263
QY 647 LVDSSGNITPTTYNISVKTTLKNSNGTSGNNKFSVNAHDN-----NSLVTAQDL---ADYL 700
Db 2264 LLNGDGLLSQRLVDV-----TLSGALDNOGSGALVSEESLTVKADQV 2307
QY 701 NKVNETADSALPSFKVQNGDNSNNAITVCKDNGKTFNTLKLKGVNGVNIITNRTGTVT 760


```

Db 2308 NNOAGTFESSAGSLVTSRGEELNNOGRLVTD--GATLNSTGFDNSRAGLVSAKGAVALRT 2366
Qy 761 FGIDQS-----NGLTTPKLVGSDTNGNRLVIRQVPSADGNSTKNIIG--LSPTLPSTIAS 814
Db 2367 GALNNSQKSGIGGNTGVTIVAGLVNDRGREGRISTKTGLDAN-----LKGLLOQGGGSLVG 2421
Qy 815 PSGRNIAL--GNTIEBKDKSNAASIDDDVLNAGFNKNN-----GKDKDFSVYDVTVD--- 864
Db 2422 ERGVTLDLNGTLDNHDGLVSTPGALLRQLGMVNSVSGEISSRAFTTLAANTLNNOG 2481
Qy 865 --FIDGNATATV--TYDEANOTSKVAYD--VNVDEKTIELTDGNGKQKLGKVTIKLTETS 919
Db 2482 CRLISSEALTURIKAKTLDNLSKQVATDGLAIESQVL-----DNRAGTIGSK----- 2529
Qy 920 TNGNATTFSTDDHALVKASDIAGNLTAEIHTTKGTANTALQTFTVKVVDENDKADD 979
Db 2530 --GDARISVTSLDNAEGSLVSEGRLELVADQV-----VNTL-----KLCKKNGLDIKTKDKGTVTFTGINTQSG 1024
Qy 980 TNAITVGKDTSGK-----VNTL-----KLCKKNGLDIKTKDKGTVTFTGINTQSG 1024
Db 2561 -----SNGNOGRTAARGVLEAVGTLLOQGGELVSGQSLDLRAD---TLD---NSQSG 2607
Qy 1025 LKAGDSTLLNNGLSIK-----NTASNEQIQVGADGVKFMVWNG--VVGAGIDGT 1073
Db 2608 LIAA-----NGGIAEARQVDNRAGEISSTSKVAVNA---REQLDNRGGKVIG---DSG 2655
Qy 1074 TRITREIGFTGNSLDKSPHLS--KGINAGGKKITNIOSGEI--AKNSHDAVTTGKIY 1131
Db 2656 LRLTVQRL-----LNOAKGVLAGRDLGSLDGELFNGGGRLDSONLSVSLGGVLD 2707
Qy 1132 DLKTELENKISSTAKTAQ-----NSLHEFSVADEQGNFT-----VSNPYSSY 1174
Db 2708 NOGGALVSEGSITARAARLDNRGGTFSSAGALATQSAVLNQGGRLLSDAGVTLKGASL 2767
Qy 1175 DTSKTSVDTTFAGENGITTKV-----NKGVRY-----GIDQTKGLTTPK 1214
Db 2768 DNSR--SGVISAKGAVDIRTVGLDNRNGGIGSGNAGITILVAARLDNNGOQGRVSAKGLLDAN 2826
Qy 1215 LTVGNNGKGVINSQNGQNTTIGLSNTLAN-----VTNDKGSVRTTEQGNII 1262
Db 2827 LKGLDORGGLVSEGTGVTDLNG--GTLVNRDGGLIATPGALLRQLGAVDNGAGGEIS 2884
Qy 1263 KDBDKTRAASIVDVLSAGNIOGN-----CEAVDFVSTYDTVNFANGNTTTAKVTYDD 1315
Db 2885 SDRAFTLAAASLD--NRGRLIGADSLTLRIAQALD--NSLAGVISGAAG--LDIAARLDN 2940
Qy 1316 TSK---TSKVYVDVND---DTTIE--VKDKKLGKVTITLTSTGTG-----ANKPA 1358
Db 2941 SAKGTLASRAGIDRVDGALDNHAEGTVSGARLTASASLDNSGKLLSGNAGLSVATCA 3000
Qy 1359 LSNQATGDAL-----VKASDIVAHLNTLSGDIQTAKGASQANNSAGYVDAGNKVI--- 1409
Db 3001 LDNAEGGOLISQGVLDVSSADLDNRGALSGKOSLSAANLDNRGGLTSDGELELTAG 3060
Qy 1410 -YDSTDNKYQAKND-----GTVDKTKVAKDKLVAQAQTPDGTILAQMNVKSV 1456
Db 3061 RVDSADGGEISARGDRLTLTVRQGRIGRGVSLDRGGDDLNOGGLISARGPLSI 3120
Qy 1457 INKEQVNDANKQGINEDNAFAVLEKAAS--DNKTKNAAVTVGDLNVAQAQ----- 1506
Db 3121 ---ERLNVLDNRQ--GEIYSQOCFELLARRIDNNGOQGRILISAGKRLRLDADALGNAGAGL 3175
Qy 1507 -----PLTFAGDT-----GTTAKKLE-----TLTIKGOTDTNKLTDN 1540
Db 3176 LSGWQGLTVTGGSLDNSAGGTLSSKGDELALISLGGALDNHIGQALYSKGAQRIDAASLDN 3235
Qy 1541 NIGVAGTGTFTVKLAKDLTLNLSVNAAGTKIDEKGISFV-----DANGQAKANTPVL 1593
Db 3236 AQGIVSGESDVTLSIAGKLDN---GQGLVSAQRALESFERRDDTLNNAAGRINGGSL 3291
Qy 1594 SANGLD--LGGKVIN-----VKGTKRDTDAANYQQLNEVRNLLGLGNDNADGNQVN--I 1644
Db 3292 KGASLDNSDGLISQGRLDATIGGALVNAAGARLASGGDL--LRSASVDNRGGKLYSQGL 3350

```

```

Qy 1645 ADIKKDPNSGSSNRRTVIKACTVL---GGKGNNDTEKL---ATGGVOVGVDKDGANGDL 1698
Db 3351 LEISAGSLDNASGTLASQADMSLRLLGGGALRNQODGLIFSQAGALEVQAGSLDNRQGLT 3410
Qy 1699 SNVWVKQKQSGKALLATYNAAGQTNVVTNNPAFAIDRINEQOIRFFHVNDGNOBPVQ 1758
Db 3411 -----QAQCDNRLRIGGALDNOAGRLDSRAGN-----LDLOSGSLD---N 3447
Qy 1759 GRNGIDSSASGKHSVAIGFQAKADG-----EAAVAIGRQTQAGNOSIAIGDNAQATGD 1811
Db 3448 GAGGVNLSAKGWLKLVGLFDNSAGVTQASLEIRAGOGVNRNQQOCHLSALGDNRIVTAD 3507
Qy 1812 -----QSIAGTGNVW--AGK----- 1825
Db 3508 FNOQGGGLYASGLLSLDGQRFNLNOGAAAGQGGKVGAGRIDFSLAGALANRFGQLESESEL 3567
Qy 1826 -----HSG-----AIGDPSTVK-----ADNSYSV--GNNNQFTDATQTDVFGVGNNTTV 1867
Db 3568 HLRAAIDNNSGSLRALGRSGSTRVLVAGDLNNAYGVLESANQDLQLGSLANAGRLIH 3627
Qy 1868 TESNSVALGSNSAISAG-----THAGTOAKKSDGTAGTTTATAGATGVKGFAGQTAVGAVS 1923
Db 3628 TGNFTGLDQGOVIRAGGELTTNGLLDIRASEWTNSSVLQAGRLNLDIGTFQTAEGKLL 3687
Qy 1924 VGASGAERRIONVAAGEVVSATSTDAVNGS---OLYKATQSIANATNELDHRHONENKAN 1980
Db 3688 AVQSFTGR-----GGDWSNDGILLASNGSLRLESGGYRGNGRATSLGDFAL---NAAS 3737
Qy 1981 AGISSAMAMASPMQAVIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGQWVFKINGSADTO 2040
Db 3738 LDLCNAASLA-----GGANVTLGAGNLLVNRGRITAAAGDLV-----ASAASL 3779
Qy 2041 GHVGAAGVAG 2050
Db 3780 NNYGTILGGG 3789

```

Search completed: September 13, 2001, 12:41:04
Job time: 32730 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:52:48 ; Search time 47.77 Seconds
(without alignments)
1472.190 Million cell updates/sec

Title: us-09-361-619-9
Perfect score: 10356
Sequence: 1 MNHLYKVFENKATGTFMAVA.....NGSADTQGHVGAAGAGGHHF 2053

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	608	5.9	2249	1 190K_RICRI	P15921 rickettsia
2	496	4.8	2003	1 YDBA_ECOLI	P33666 escherichia
3	456.5	4.4	1643	1 OMPB_RICPR	Q53020 r outer mem
4	455.5	4.4	1656	1 OMPB_RICJA	O08653 r outer mem
5	453.5	4.4	1577	1 HLYA_PROMI	P16466 proteus mir
6	443.5	4.3	1654	1 OMPB_RICRI	Q53047 r outer mem
7	439.5	4.2	1902	1 P2P_LACLA	P15293 lactococcus
8	436.5	4.2	1645	1 OMPB_RICTV	P96989 r outer mem
9	432.5	4.2	2334	1 WAPA_BACSU	Q07833 bacillus su
10	423.5	4.1	1902	1 P1P_LACLC	P16271 lactococcus
11	423.5	4.1	1902	1 P3P_LACLC	P15292 lactococcus
12	410.5	4.0	1848	1 CBPA_CLOCL	P38058 clostridium
13	404	3.9	1608	1 HLYA_SERMA	P15320 serratia ma
14	404	3.9	3591	1 FHAB_BORPE	P12255 bordetella
15	399	3.9	1902	1 P2P_LACPA	Q02470 lactobacill
16	373	3.6	1300	1 120K_RICRI	P14914 rickettsia
17	373	3.6	1325	1 YDEK_ECOLI	P32051 escherichia
18	366	3.5	1569	1 YFJA_ECOLI	P52143 escherichia
19	355	3.4	1286	1 AIDA_ECOLI	Q53055 escherichia
20	345	3.3	1183	1 CNA_STAAU	Q53654 staphylococ
21	339	3.3	1694	1 IGNO_HAEIN	P49696 haemophilus
22	337.5	3.3	2628	1 HAGA_FORGI	Q51845 porphyromon
23	335	3.2	1702	1 IGAA_HAEIN	P45384 haemophilus
24	334.5	3.2	1419	1 ALAL_CANAL	O13368 candida alb
25	334	3.2	1567	1 ICEN_XANCT	P18127 xanthomonas
26	334	3.2	1829	1 FRPC_NEIMC	P55127 neisseria m
27	330	3.2	1176	1 SLAP_BACSH	P38537 bacillus sp
28	330	3.2	1849	1 IGAA_HAEIN	P45386 haemophilus
29	327	3.2	1829	1 FRPC_NEIMB	Q9JYV5 neisseria m
30	315	3.0	1025	1 SLAP_CAUCR	P35828 caulobacter
31	313	3.0	1541	1 IGAL_HAEIN	P42782 haemophilus
32	312.5	3.0	1256	1 MRP_STRSU	P32653 streptococ
33	312	3.0	1322	1 ICEA_PANAN	P20469 pantoea ana

ALIGNMENTS

RESULT 1

```

190K_RICRI 1
ID 190K_RICRI STANDARD; PRT: 2249 AA.
AC P15921;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
   repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990).
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -!- PTM: GLYCOSYLATED (POSSIBLE).
CC -!- DISEASE: PROBABLY ROCKY MOUNTAIN SPOTTED FEVER.
CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M31227; AAA26380.1; -
PIR; A41477; A41477.
Antigen; Repeat; Signal; S-layer: Glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 2249
FT DOMAIN 212 1180
FT REPEAT 212 286
FT REPEAT 287 358
FT REPEAT 359 430
FT REPEAT 431 505
FT REPEAT 506 577
FT REPEAT 578 652
FT REPEAT 653 724
FT REPEAT 725 799
FT REPEAT 800 874
FT REPEAT 875 949
FT REPEAT 950 1021
FT REPEAT 1022 1093
FT REPEAT 1094 1165
FT REPEAT 1166 1180
FT REPEAT TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

```

```

34 311 3.0 918 1 YMTB_CAEEL P34487 caenorhabdi
35 306.5 3.0 1592 1 GTF2_STRDO P27470 streptococ
36 305 2.9 1258 1 ICEN_ERWHE P16239 erwinia her
37 297.5 2.9 1039 1 AGA3_ECOLI P39180 escherichia
38 297 2.9 1545 1 IGA3_HAEIN P45385 haemophilus
39 295 2.8 1228 1 SLAP_BACST P35825 bacillus st
40 295 2.8 1341 1 VG37_BPT2 P07067 bacterioph
41 292 2.8 2491 1 TALA_DICDI P54633 dictyostell
42 291.5 2.8 3562 1 PGCV_CHICK Q09053 gallus gall
43 284.5 2.7 933 1 SLAP_CAMFE P35927 campylobact
44 284 2.7 1537 1 FLOI_YEAST P32768 saccharomyc
45 284 2.7 1861 1 APU_THETU P38536 t anylopull

```


Thu Sep 13 14:19:31 2001

Query Match 5.9%; Score 608; DB 1; Length 2249;
 Best Local Similarity 22.5%; Pred. No. 1.1e-14;
 Matches 491; Conservative 224; Mismatches 765; Indels 704; Gaps 114;

82 INNTLKDALATGEASTAFGSLKAGOSQIAIGSVKPDPNNGSNGVSH-----132
 13 IQOGLKAALFTTAAIMLS-----SGALGVATGVIATNNAAFSNNVNNMNEITAAAG 68
 133 -----AKGNESIAIGDVLAEAGDASTAI-----GSDDLYLPKMLDKNEFHKL 176
 69 VANGTPAGGPQNNWFTYGGDYTTADAADRIIKAINVAGT-----TPVGLNI-----116
 177 HGEHILKIQISTGKIKYR-----TRAOGHASTAVGAMSQAQGHFSNA-----FT 224
 117 -----TONTVVGSIITKNLLPVLNAGKSLTLNGNNAVAANHFADPADNYTGLGN 168
 225 YATAEAYSLAVGLAAQATKOSSIAVGSN-----AKANAFATAGTNT-----VVNL 272
 169 IALGGANAALII-----QSAAPSKITLAGIDGGIITVKTDA-AINTGICNTNALATVNV 223
 273 GRGVALGFGSQILDRDN--TDASAVPL--GKTLADQYKATROGSDTDFSI-----321
 224 GAGTATLGAVIKATTTKLTNAASVLTLTNANAVLTGNAIDNTTGGDNVGLNLGALSQV 283
 322 -GNSNNNSIRRKIIINVAGSRDID--AVNVAQKLVEELANKRIT-----FKGDGDNNS 374
 284 TGDIGTNS--LATISVAGATATLGAVIKATTTKLTDAASAVKFTNPVVVTAIDNTG 340
 375 NSVERGL--GNTLIKDAOTNALTENIG--VVTGNGCLKVAKELTGLTSVSAT 427
 341 NA--NNGIVFTGNTSTVGVNGTINATVNVGAGLQVGVVKNANTINLTNDASAVTET 399
 428 NKITVSTNNNAELQSGILTF--SPIGTKTDTKYVSIDGLKFTNDSNIATKG-----480
 400 NPVVVTAIDNTGNANNGIVFTGNTSTVGT--DIGTINATVNVGAGTATLGAVI 454
 481 ---TTRITKKKIGFAGTNDGVDESKPYLDNEKLKVGNSLTNSGSLTVNNTTKNQIQV-G 536
 455 KATTTKLTNAASVLTLTN-----ANAVLTG--AIDNTTGGDNVGLN 494
 537 ANGKFKATVANNVANT-----SATVG-----TARITEKIGFAGTNDGVDEQ 578
 495 LNG-ALSOVTGNTGNTSLATISVAGATATLGAVIKATTTKLTDAASAVKFTNPVV--550
 579 APYLDKERLKVREIIR--TDSGT--NAGNHKIT--GLTNGIAN-----616
 551 -----VTGAIIDNTGNANNGIVFTGNTSTVGTGDTGNTNSLATISVAGATATLGAVI 601
 617 -----TDAVITIKOLKDAKPTL-----NAGDISINNSNGDLVDSGNI--TPTTYN 660
 602 KATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNLGALSQVTDIGTNTSLAT 661
 661 ISVKTTLKNSGTSNKNFVSNAHDNNSLVAKDLADYLNKYNETADSALPSEKQVONGD 720
 662 ISVGA-----GTATLGAVIKATTTKITNAVASVK-----FTNPVVVVG-----AIDSTG 706
 721 NSNNAI---TVGKDRNGKFTNTKLKLGENVNITNATGTVTIGDQSGN-----LTPK 773
 707 NANNGIVTGTGNTSTVGTGDTGNTAL-----ATVNVGAGTATLG--GAVIKATTTK 754
 774 LTVGSD-----TNGNRLVIEQVPSADGNSNTKNIK-----GLSPPLPSIASPSG 817
 755 LTNAASVLTLTNANAVLTGAIDNTTGGDNVGLNLGALSQVTDIGTNTSLATISVAG 814
 818 RNIALGNTIEEDK--SNAASIDVNLNAGFNKNGKDKDFSVYDVFIDGNATATV 875
 815 TATLGAVIKATTTKLTNAASVLTLTNANAVLTG--AVDNTTGGDNVGLNLGALSQV 871
 876 TYDEANOTSKVADVNVDEKTIETLTGDNKKOLGVKTIKLT-----TSTGNATFTSTD 930
 872 TGDIGTNTSLATISVAGATATL-----GGAVIKATTTKLTNAASVLTLTNANAVLTGAI 925
 931 D-----DHALVKASDIAGNLNLTAEIHTTKGTA---NTALQITVTKVKVDEN 974

926 DNTTGGDNVGLNLGALSQVTDIGTNTSLA--TISVAGATATLGAVIKATTTKLTDA 984
 975 DKADDTNATV--GKDTSGKVNLTCLKKNGLDIKDKDGTVTGINTOSGLKAGDSTTL 1033
 985 SAVKFTNPVVVTAIDNTGNAN-----NGIVTF-----TGNSTVT 1019
 1034 NNGLSITKNTASNEQIQVGADGVKFAWNVNVVVGAGIDGTTRITRDEIGFTGTN--G 1088
 1020 GNVG-----NTNALATVNVGAG--LLQVQGVVKA--NTINLTNDASAVTTPVNVVG 1069
 1089 SLDKSKPHLSKDGGINAGKKITNIQSGEIAKNSHDAVTGKIYDLKTELENKISSTAKTA 1148
 1070 AID-----NTG-----NANNGIVTFTGNTSTVGT-----NVGNT-----1097
 1149 QNSLHESVA--DEGNNFTVSNPYSSYDTSKTSVITFAGENGITTKVNGVVRVGI 1204
 1098 -NALATVNVGAGLQVQGVVKA--NTINLTNDASAVTF-----TNPVVVTAI 1143
 1205 DOTKGLTTPKLTGVN--NNGKIVINSONGONTIING--LSNTLANVTNDKGSVRTTEOGNII 1262
 1144 DNT-----GNANG--IVTFTGNTSTVGTGIDGNTNALATVNVGAGITLQAGSL 1189
 1263 KDEDTRAASIVD-----VLSAGENLOGNEAVDFVSTYDTPNFANGT-----1306
 1190 -----AANNIDFGARSTLBPENGLDGGKRAIPY--YFKGAIANGNNAITLVNTKLLT 1239
 1307 ---TTAKTYDDTSKTSKVYDVNVDDTTI-----EYDKKLGKVTTLTSTGTGA 1354
 1240 ASHLTIGTVAEINAGNLFTIDASVGDVTTILNAQINFRARDSVVLVLSNL--TGCV 1295
 1355 NKFPALNQATGDALVKASDIVAHLNLTLSGDIOT-----AKGASQANNASGVV--DAD 1404
 1296 NNTLL-----AADLVAP--GADEGTVVENGVLNGLVNSVAGTARRIGDGG 1340
 1405 GNR-----VIVDS--TDNKYQAKNDGTVDKTEKAVAK-----DKLVA 1439
 1341 GNFENTLLIYNAVITITDDVNLEGIONVLINKNADFTSSTAFNAGAIQINDATYTTDANNG 1400
 1440 QAOTPDGTL--AQMNVKSVINKQVNDANKKQINEDNFAVKGLEKAASDNKTKNAVTV 1497
 1401 NLNIPAGNIQFADADAQVLQNSSGNDRTITLCANID-----PDNDEGIVI--1447
 1498 GDLNAV--AQPLTFAGDTGT--AKLGETLIG--GOTDTNKLTDNNIGVAGTGDFT 1552
 1448 --LNSVTAGKLLTAG--GKTFGAHL--QTLFLKAGDCST-----AGTPTT 1491
 1553 VKLAKDLTLNLSVNAVGTGKIDKGISEVDANGAKANTPVLSANGLDLGGKVISNVKGT 1612
 1492 TNVLDTI-----GO-----LELG-----1505
 1613 KDTDAANY-----OQLNEVRNLLGLGNDNAD-----GNOVNIADIKKOPNSGSSNRT 1660
 1506 --ATTANVLFENDAVQLTQTGNIGGFLDFNAKNMVTLNNVNVAGAVQ--NTGGTNGT 1561
 1661 VIKAGTVLGGKGNNDTEKLA--TGGVQGVVDKDGDNAN--GDSLNWVVKTKDSSKALLA 1716
 1562 LI-----VLGASHNLKRVNGIAMLKVGAGNVITAKGKVKIGIEGTGTNT-----LTPA 1611
 1717 TYNAAGOTNYVTNNPAEALDRINEQGIREFHVNDGNOEPVQGRNGIDSSASGRKHSVAIG 1776
 1612 HENLTG-----SINKTGGQALKLNFPMNGSVSCVV-----1641
 1777 FQKADGEAAVAIGROTQAGNOSIAGDNAQAT-----GDQSI-----IGTGNVWAGKUSGA 1829
 1642 -----GTAAHSVGDITTAGATSPASSYNAKCTATLGGTTFANTFTTGAVTLAK-----1691
 1830 IGDPSVTKADNSYSVGNNOQFDTATQTDVFGVGNNTVTESNSVALGSNSAISAGTHAGT 1889
 1692 -----GSITSFAKNVTATSF--VANSATINFSLSAFNSN--ITGG-----1728
 1890 QAKKSDGTAGTATTTTGA-----TCTVKFGAGQTAAGVAVSVA--SGAERRIONVAAGEVS 1942

Db 1729 -----GTTTLGCAQVYTGCT-GSFTDTLTNTTDFGAAGKSGNLLIKSGSTLDLS 1778
QY 1943 ATSTDAVNGSQLYKATQSTANATN 1966
Db 1779 GVSTLAL-----VVTATN 1791

RESULT 2
YDBA_ECOLI
ID YDBA_ECOLI STANDARD; PRT: 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Kasei H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Aiba H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimura H., Nishio Y., Oshima T., Saiko N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RN SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000237; AAC74483.1; ALT_SEQ.
DR EMBL; AE000237; AAC74487.1; ALT_SEQ.
DR EMBL; D90778; BAAL5009.1; ALT_SEQ.
DR EMBL; D90778; BAAL8880.1; ALT_SEQ.
DR EMBL; D90779; BAA18881.1; ALT_SEQ.
DR EMBL; X62680; -; NOT_ANNOTATED_CDS.
DR EcoGene; EG11307; ydbA.
KW Hypothetical protein.
FT CONFLICT 489 489 I -> V (IN REF. 2).

FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.8%; Score 496; DB 1; Length 2003;
Best-Local Similarity 20.6%; Pred. No. 8.9e-11;
Matches 399; Conservative 227; Mismatches 663; Indels 646; Gaps 99;

QY 284 ILDRNNTDASVYPLGKTLADQ-YKATRGQDSTDFISGNSNNSSIRKINVGAGS 342
Db 147 ILDKTEKT-----LTIKRSVFYVTENADGT--ISLQDSNGRKATI----- 184
QY 343 RDTDAVNVAAQLKVLVEELANRKITFKGDCDNNNSVERGLGNTLTIKGDAQTALTEANIG 402
Db 185 -----NLWQI-----DEANNTVALEGSADGATKWQYNHNGELVITGD---NATVNNKG 231
QY 403 VYTDGNGLVKLAKELTGLTSVSATNKITVNTNNAELOSGLTFSPTGKTDKTVY 462
Db 232 TVVDG-----KDSGTG-TEIN-----GNKGVIQDGDLDVSG----- 261
QY 463 SIDGLKFTNDSNSIATKGTTRIT-KKIGFAGTGDGDESKPYLDNEKLVGNSTLNSGS 521
Db 262 GGHGIDITGDSATVDNKGTMVTPDPESMGIO-----IDGKAIVNNE---GESTINGG 312
QY 522 LTVNNTGKQIQVANGANGIKFATVANNAVNTSATVGTARITEEKIGFAGTNDGVDQAPY 581
Db 313 -----TGQT-----INGDD--ATANNGKT-----TVDGKDSGTGE 341
QY 582 LDKERLKGVRVEITDTSGINAGNH--KITGLTNGIANTDAVTIKQLKDAKPTLNAGD--- 636
Db 342 INGN--GKVIQDGDLDVSGGGHGIDITG-----DSATV-----DNKGTMTVTPDES 386
QY 637 -GISIN-----SNGDLVDSSGNIITPTVYINISVKTKLN-SNGTSGNNKFSVSNADNN 688
Db 387 IGIVQDGDQAVVNEGESAITNG-----TGQINGDDATANNGKTTVDGKDS- 435
QY 689 SLVTAKDLADLYNKVNETADALSALPQVQNGDSNNNAITVKGDTNCKTFTNKLKGENGV 748
Db 436 ---TGTEIAGNGKV-----IQDGD-----LDVSG-----GGHGI 462
QY 749 NITNTRAT---GTVTFGIDQSNGL-----TTPKLTGSDPTNGNRLVIEQ 789
Db 463 DITGDSATVDNKGTMVTPDPESIGIQIDGQALVNGEGESTITNGGTGTGIN- 516
QY 790 VPSADGNSTNIIKGLSPTLPSPSPG-----RNIALGNTIEERDK- 831
Db 517 -DATANNSGKTTVDGKDSGTGKIAGNIGIVNLGSLTVTGAGHVENIGDNGTVNNKGI 575
QY 832 --SNAASIDDLVNA-CFNLKNNCKDKDFVSTYDTVDFIDGNATATVTVYDEANQTSKAY 888
Db 576 VVSDTGSIGVLINGEGATVSNITGDVN-----VSNEATGFSITTSNGKSVSLAGSM 624
QY 889 DVNVDEKTELTGDNCKKOLGVKTIKLTETSTNGNATFTSTDDHALVKASDIAGNLNTL 948
Db 625 QVGDPTGVDLNGNNSVTLAAKDLKVVQKATG-----INVSGDANTVNIT 671
QY 949 ABEIHTTKTANTALQFTTVKKYVDENDKADDTNATVKG-----DGTSGKVTNLTCLKG 1001
Db 672 GNVLVKDKTADNAAEYFFDPSPGVINVYSDNNVTLDGKLTVVSDSEVTSRQSNLF----- 727
QY 1002 KNCGLDIKTDKDTGTVTGINTQSGLKA-GDSTTLNNG-LSIKNT-ASNEQIQVCAAG 1055
Db 728 -----DGSA-----EKTSGLVVIGDNTVNMGGLELIGEKNALADQSVTSILRTG 773
QY 1056 VKF--AMVNVNGVVGAGIDGTTTRITRD-EIGFTGTNGSLDKSPHLSKDGINAGKKTINI 1112
Db 774 YSXTSVIVVSGESSVYVNGDFTISGEPPLGAGV-----IRVDKALLEI 818
QY 1113 QSGEIAKNSHDAVTKGKIYDLKTELENKISSTAKTQNSLHEFSVADEQNNFTVSNPYS 1172
Db 819 GSG-----ATLTMQDIDDSFEHHG----- 836
QY 1173 SYDTSKTSVDV--ITFA---GENGITTGVNKGVVVRVIGDQTKGLTTPKLTGVNNGNK---- 1223


```

Db      837  -----TRTVEIQNLGEFAFVTGEN--TTGINSGLTSLONGKDPAPSPVILLATNGSGATNA 890
QY      1224  GIVINSQNGNTI-----TGLSNTLANVTNDKGSVR--TTEQGNIIKDEKTRAASIVDYL 1277
Db      891  GTITGKVTBQHSVENKYSTGTSNSFI--FNNDVSSITCLVAQSN-----STIINTD 939
QY      1278  SAGNLOGNGEAVDFSTYDVFNFANGNTTAKVTYD--DTSKTSKVVDVNVDDTTIEV 1335
Db      940  SGIIDLYGRG--SVGMLAIADSTAENOGKITLDSMWVDANDTTAMRDIAANSATDFGT--- 995
QY      1336  KDKLGVKTTTLTSTGTCANKFALSNGATDALVKASDIIVAHNLNTLSGDIQTAKGASQAN 1395
Db      996  -----GVGVGSDTSVGAGKATATL--NLGG-----VITIYNAGAG--MAAYGAS--- 1036
QY      1396  NSAGYVDADGNKVIYDSTDKYIQAQKNDGTVDKTKVEKAKDKLVAQAQTPDGTGLAQMNVKS 1455
Db      1037  -----NTVINQGTIN-----LEKNGYDGS--LAANTLVGMAYVEHGT----- 1072
QY      1456  VINKEQVNDANKKOGINEDNAFVKGLEKAAASDNKTKNAATVVGDL-----NA 1502
Db      1073  -----AINDQGTGVININVGT-----QAFYNDGTGTIVNYGTICTFGVCQSGNEYNNT 1120
QY      1503  VAQPLTFAGDTGTTAKKLGFTLTKGQDQTDNKLTON--NIGVYAGTGTFTVKLAKDLT 1560
Db      1121  DDFTSLIYTG--GDTITRSGETVTLNKSAAVTDKLAGNVVNSGTLSG--DQITVS----- 1172
QY      1561  NLNSNVNAGG-----TKDEKISFVDANGQAKANTPVLNGLDLGKGVISNYGK----- 1610
Db      1173  GILLETSGGIIINLVKLDK-----GAVIKNAGVMT--NNVDVSGGILLNAGEMTAQI 1222
QY      1611  ---CTKQDAAVQQLNEVRNLLGLGNDNADGNQVNIADIKKDPNSGSSNRTRVIKAGT 1666
Db      1223  TMAAGADSLVNYNTGINKIVQNAVFN-----NSGSVTGRMW----- 1260
QY      1667  VLGGKGNNDTEKLTAGGVQGVQVDKNGNANGDLSNVW-VKTKQDKSKKALLATYNAAGOTN 1725
Db      1261  SAGGVFNQTD-----GAIMRGAALGTGTAVANNEGTWNILGSSSEGNNTGMLEVNNSAFNN 1316
QY      1726  ---YVTNPAEAIIDRINQGITRFFHVNDGNOEPVYVQGRNGIDSSASGKHSVAIGFQAKAD 1782
Db      1317  REGFILNDKNAV--HINSGTLY---NTGHM-----NLSNHHGAVNMW----- 1357
QY      1783  GEAAVAIGRQTOAGNOSTAIGNAQAOTGOSTAIGTGNVAVACKHSGALGDFSTVKADNSY 1842
Db      1358  -----GGNGRFINDGTIDVSAKSLV-----VSANNA- 1383
QY      1843  SVGNNOFTDATQTDVFGVGN-----ITVTSNSVALGNSAISAGTHAGTQAOKSD--- 1895
Db      1384  --GDQNAFFWQDNGVINFDHDSASAVKTHSNFTA--QNDGIMNISGTGAVAMEGDKNA 1439
QY      1896  -----GTAGTFTT-----AGAT-----GTVKFGAGQTAVGAVSVGASGA 1929
Db      1440  QLVNNGTINLGTAGTDTGCMIGMOLDANATADAVIENNGTINIFANDSFASFV-LGTVGH 1498
QY      1930  ERLIQNVAAGEVSATSDAVNSOLYKATQSI-----ANATNLDHRIHONKANAGISS 1985
Db      1499  -----VYNNGFVWADGTGVSGLLKQGDGINVEGMNGNNSSEVHGD----- 1542
QY      1986  AMAMASMPQAIYIPCRSMVVTGGI-----ATHNGQCAVAVGLSKLSDNGQWVKFKNGSA--- 2037
Db      1543  -----YTLPDVPPKPTVNTVSGDSAGGSMNLLNGYVVG'N-----VNGSAGKLIK 1587
QY      2038  -DTQGHVGAAGVAGCF 2051
Db      1588  VNNASMGVBEINTGF 1602

```

RESULT 3

OMPBL_RICPR

ID OMPBL_RICPR STANDARD; PRT: 1643 AA.

AC Q53020; Q92CM0;

DT 30-MAY-2000 (Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_taxid=782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RX MEDLINE=91045972; PubMed=2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BREINL;
RX Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompB of Rickettsia prowazekii.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [4]
RP PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104669; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplik W. Jr., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
RN [6]
RP STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
RP VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
RN [7]
RP FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
RN [8]
RP SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
RP LAYER WITH HEXAGONAL SYMMETRY.
RN [9]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@sib-sib.ch).
RN [10]
RN EMBL; M37647; AAA26390.1; ALT_INIT.
RN EMBL; AF161079; AAD42294.1; -.
RN EMBL; AJ235273; CAA15140.1; -.
RN Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
RN CHAIN 1 1328
FT

```



```
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 Y -> A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
FT CONFLICT 178 179 AA -> VC (IN REF. 1).
FT CONFLICT 191 201 TQEAPLTLGA -> INSRSSSYHLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735PDE392E6346CC CRC64;

Query Match 4.4%; Score 456.5; DB 1; Length 1643;
Best Local Similarity 22.4%; Pred. No. 1.8e-09;
Matches 396; Conservative 227; Mismatches 628; Indels 517; Gaps 100;

QY 517 LNSGSLTVNNTGNKIOVGANGIKP--ATVANNVANTSATVGTARITEKIGF---AGT 571
DB 12 ISAGLVTAATAT-----IVAGFGVGAAMAOYNRITNAAT-----TFDGGFDQAAGA 61
QY 572 NPGVDQAPYLDKERLKVGRVEITDSDG----INAGNHKIKITGLINGIANTDAVTI----- 622
DB 62 NIPV---AP-----NSVITANANNPITFNTPNGLNSLFLDTANDLAVTINEDTT 108
QY 623 -----KOLDKAKPTLNAAGDISINSGDLVDSGNIPTTPYINISVKTKTLNSNGTS 674
DB 109 LGFITNIAQAQAFNFNTVAAGKILNI-TQGGITVQEASN---TINAQNALTKVH----- 158
QY 675 GNNKSVSNAHDNNSLVATKADLADYLNKYNETADSLPSFKVONGDNSNNAITVGKDTNG 734
DB 159 -----GGAINANDLSG-LGSITFAAPSVLEFNLINPTQEAFLTG---ANS 203
QY 735 KTFNTLKLKXENG-----VNITNTRATGVTFGIDQSNGL---TTP-----KLTVG 777
DB 204 KIVN-----GGNGTLNITNGFIQVSDNTFAGIKTINIDCQGLMFNSTPDAANTLNLOVG 258
QY 778 SDT---NG-----NRLVIEQVPADGNSTK-NIIKGLSPILPSIASPSGRNIALGNITIEK 829
DB 259 GNTINFNGDGTGKLVL---VSKNGAATFENVTGTLGGNLKGIELNTAAVA-GKLISQG 314
QY 830 DKSNAASIDDLVNLNAGPNLKNKGKDFEVTYDVFDFIDGNATTTAT-----VTYDEANQ 882
DB 315 GAANAVIGTD-----NGAGRAAGFIVSD-----NGNAATISQVYAKNKNVIOSANA 361
QY 883 TSKVAYDVNVDEKTIELTDGNCKQLQGVKTIKLTETSTNGNATFTSTDDHALVKASDI- 941
DB 362 GQOVTEHIVD---VGLGCTTFKTAADSKVI-ITENS-NFGSTNFGNLDTQIVVPDKIL 416
QY 942 AGNLNTLAEIHTTKTANTA-LOFTVK-----KVDENDKADDTNATIVGKDGTF---S 991
DB 417 KGNF-----IGDVKNNGNTAGVITFANGALVSASTDPNIAVTNINATIEAEGAGVVELS 470
QY 992 GKVNTLKLKGNCLDIKTKDQGVTFVFGINTQSLKAGDSTTLNNGLSIKNTASNEQIOV 1051
DB 471 G-IHIAELRLNGGGSIFKLADGTIVNG-----PVQN----- 501
QY 1052 GADGVKAFMNVNNGVWAG---IDGTRITRDEIGFTGNSL-----DKSKPHLSKDG 1101
DB 502 -----ALMNNNALAAGSIQLDGSAILITGD-IGNGVVNAALQHIHLANDASK-ILALDG 552
QY 1102 INAGGKKIYNISQGETAKNSHDAVTGCKI-----YDLKTELENKISSAKT 1147
DB 553 ANIGANV-----GGAT-----HFQANGTKIKLTNTQNNIVVNFDFDITTDKTVGVVDASSLT 604
QY 1148 AQNSLH-EFSVADEQGNFTVSNPYSSYDTSKTSIDVITFAGENGITTKVKNGVVRVGIQ 1206
DB 605 NQOTLINGSIGTVVANTKTLAQ--UNIGSSKT---ILNAGDVAINELVNIENNGSVOLNH 659
QY 1207 TKGLATPKLTGVNNGNGKGVINSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDED 1266
DB 660 NTYLIITKTINAAN---QGQIIIVAADPLNTNT---FLADGTN-LGSA-----ENPLSTIHF 707
```

RESULT 4

```
OMP_RICJA ID OMP_RICJA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (166 KDA SURFACE-LAYER PROTEIN)
```

```
QY 1267 KTRAAISIVDVLSA--GFNLOGNGEAVDFVSTYD-----TVNFANGNTTTAKVTV--DDTSK 1318
DB 708 ATKAAADSLILNVGKGVNLYANN-----ITNDANVGSLSHFRSGGTSIVSGTGGQOCHK 762
QY 1319 TSKVVYDVNVDDTTIEVKDKKLGKVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHL 1378
DB 763 LNNLLILD---NGTTV---KFLGDTTFNGGKTIEGKLSILQISNNYTTDH-VESADNTGTL 814
QY 1379 NTLSGDIOTAKGASQANNSAGY-----VDADGNKVIYDSTDNDKYYQAKNDGTVDKTK 1430
DB 815 EFVNTDPIIV---TLNKQAGFGLVKQVIISGPN-IVFNEI-----GNVGIVH 859
QY 1431 EVAKDKLVAAQAT-----PDGT-LAQMNVKSVINKEQVNDANKKOGINEDNAFVAGLE 1482
DB 860 GIAANSISFENASLGTSLSLPSGTPLDVLTIKSTVGNGTYDN-----FNAPIVVVSGID 913
QY 1483 KAASD-----NKTNAAVTVGDLNNAVAQTPLTFRAGDTGTTAKKLGETLTTKGGQ----- 1531
DB 914 SMINNGIIGDKKKNIILSLGSDNSITVNANTLYSGIRTTKNQO--TVTLSGGPNPNPGT 972
QY 1532 -----TDTNKLTD---NNI-----GVVAGTDGFTVKLAKDL 1559
DB 973 IYGLGLENGSPKLGQVTFFTTDYNNLSSIIANNVTINDYVTLTTGGIAGTD-FDAKI--- 1027
QY 1560 TNLNSVNAAGTKKDEKISFVDAN-----GQAKANTPVLSSANGLDLGKVISNVG 1609
DB 1028 -TLGSYNGNA-----NVREVDSTFSDPRSMIVATQANKGTVY-----LGNALVSN- 1073
QY 1610 KGFKPDAAANVOQLNEVRNLLGLGNDNADGNNOYNIADIKKDPNSGSSSNTVTKAGTVLG 1669
DB 1074 -GSLDTPVASVR-----FTGNDSGAGLQGNITYSQNIID---FGYNTILNSNVILG 1120
QY 1670 GK-----GNN-----DTEKLATGGV-QGVGVRDKG----- 1692
DB 1121 GGTFAINGEIDLTLNLIIFANGTSTWGDNTSISTTLNVSSGNIQGVVIAEDAQVNAATTG 1180
QY 1693 -----NANGDLSNVVVKTKQDKGSKALLATYNAAGQFNVTNNPAPAIIDRINQGI 1743
DB 1181 TTTIKIQDNANANFSGTQATYTLIQGG-----ARFNGTGLAPNFAVTG-----SNIFV 1227
QY 1744 RPFHVDNGNQPVPVQVRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSTAIG 1803
DB 1228 KYELIHDSNODYVTRTNDVNLVVT-----TAVGNSAIFA-- 1261
QY 1804 DNAQATGDQ-SIAIGTGNVVAGKHSIGAIGDPS-----TVKADNSYSVG-----NNNQFT 1851
DB 1262 -NAPGVSONISRCLESTNTAAYNMMLLAKDPSVATPVGAITDTSAAVTIVNLNDTQKT 1320
QY 1852 DATQTDVFGVGNITVTESNSVALGNSALSAGTHA-----GTOAK-----KSDGTA 1898
DB 1321 QDILLSNRLGLTRYLSNAETSDVAGSATGAYSSGDEAEVSYGVNAKPFYNTAEQDKKGIA 1380
QY 1899 GTTTTAGATCTVAG---FAGOTAVGAVSVGASGAERIQNVAAAGEVSATSDVAVNG--SQ 1953
DB 1381 G--YKATGTGVVGLDTLASDNLMLGAAIGITKTDIKHQDYKKGD-----KTD-INGLSFS 1433
QY 1954 LYKATOSIAN-----ATNLDHRIHQENKANAGISSAMAMASMPQAYIPGRSNVT 2004
DB 1434 LYGSQQLVKNFFAAGNAIFTLNKKVKSORSYFFESNGKMSKQIAAGNYDNMTF--GGNLIF 1492
QY 2005 GGIATHNGQCAVAV-----GLSKLSDNCO 2028
DB 1493 G--YDYNAMPNVLVTPMAGLSYLKSSNE 1518
```


DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB.
OS Rickettsia japonica.
ON Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; --
KW Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1338 32 KDA BETA PEPTIDE.
FT CHAIN 1339 1656 POLY-GLY.
FT DOMAIN 528 533
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
SQ

Query Match 4.4%; Score 455.5; DB 1; Length 1656;
Best Local Similarity 22.0%; Pred. No. 2e-09; Indels 533; Gaps 96;
Matches 397; Conservative 204; Mismatches 673;

QY 490 GFACITNGVDESKPYLDNEKLVGNSGLTVANTTG-NKQIOVGA-NGIKPATVAN 547
DB 45 GVATTVDOVG-----PDQVALANVAVAPNAVITANANNINGINLNTAGSFNGL-FLSNAN 98
QY 548 NVANTSATVGTARTEKEKIGFAGTNDGVDEQAPYLDKERLVGRVEITDSDGINAGNHKI 607
DB 99 NLAVT-----VSE-----DITLGFINNAANNANFNITLDAGKTL---TI 135
QY 608 TGLTNGTANTDAVTIKOLKDAKPTLNAGDGLISNSNNG-DLVDSSGNITPTPTNIVSKTT 666
DB 136 TG--QGITNVSQAATHNAQNIKAVFNGGAALANDLSGLGTIDFGAAASTLVDFLANPT 193
QY 667 K-----LNSNGTSGNKKFVSNAHNNLSVTAADLADYLNKVN---ETADSAALPS 713
DB 194 OKAPLILADNALIVNGANCT-----LNVNNGFIQVSDKSFAT 230
QY 714 FRVQNGDNNNAITVGK-----DTNGKTFNTLKLKNGVNIITNTRATGTVTFGIDQSN 767
DB 231 VK-----AINIGDQGGFMTNATNALNLQA-GGTTINFNCTDGTGRVLVLSKN 280
QY 768 GLTTPKLVGSTNGLNRLVIGQVPSADGNSYKNIKGLSPLTLPASPSGNIALGNTIE 827
DB 281 GAAT-----DFN-----VTGSLGNLKGIE-----LNVIAI-NGOLIA----- 313
QY 828 EXDKSNAASIDIVLNAENLKNNGDK--DFVSTYDT--VDFIDGNATFATVYDEANQT 883
DB 314 ----NAGPANAVIGT-----NNGAGRAAGFVSVSDNGKAATIDQGVYAKDMVIOQSANAN 363
QY 884 SKVAY----DVNVDEKTIELTGDNGKKQLGVKTIKLTETSGNATTTSTDDH--ALVK 937

DB 364 GQVFRHIVDVGIDGTTAFKT-----AASIVAITQNSNFG-----TTDFGNLAQVTT 410
QY 938 ASD-----TAGNLNTLAEELHTTKGTANTALQ--TFVVKVVDENDKADDTNATVGVKDGTSKG 993
DB 411 VPDMTLTGFTGDANNPNTAGVITFAANGTASADANAV--TNNTAIE--ASG-- 465
QY 994 VNTLKLKGNGLDIKTDKDGTVTFEGINTQSGLKAGDSTLLNNGLSIKNTASNEQIOVGA 1053
DB 466 VGVVOLSGTHTAELRLGNAGSV-----FKLADGTVINGK-----VNOTVILVG-- 507
QY 1054 DGVKFAMVNNVGVGAG-----IDGTRITRDEIGFTGTCTNGSLDKSPHLSKDG---INAGGK 1107
DB 508 -----GVLAAGAITLDGSATITGD-IGNGGGGAALQSII--LANDATKTLTLGGA 554
QY 1108 KTNIOSGEIAKNSHDAVGTGGKIYDLKTELENKISSTAKTAQNSL-----HEPSVADEGNN 1164
DB 555 NIIISANGGTI-----NFQANGGTI-----KLTST-----QNNIVVDCDLAIATDQTV 597
QY 1165 FTVSNPYSSYDTSKTSDDVITFAGENGITTKVNGGVVRVGIQDTK-----GLTTPKLTGVNN 1220
DB 598 VDASS-LTNAQTLLTISGTIGIIGANNITL-----GQFNIGSSKTTLLNGVAINELVIGNN 652
QY 1221 N-----GKG-IVINSONGQNTITGLSNTLANVYNDKSGVTRTEQONI 1261
DB 653 GSVQFAHNTYLTIRTNAGQKGIIFNPVNNNTTLAAGTNLGSAAANPLAEINFGSKG-- 710
QY 1262 IKDEDKTRAASIVDVLSAGFLOGNGEAVDFVSTVTVNFAN-----GNITTAKV 1311
DB 711 -----ARADTVLNV-----GEGVNLATNTITTTDANVGSFVFNAGGKNIVSGTV 754
QY 1312 TYDDTSKTSKVYDVNVVDTTIEVKDKLGVKTTTLTSTGTGANKFALSNAQATGDALVKA 1371
DB 755 GGOQGNKFNFTVALD---NGTTV---KFLGNATFNNGNTTIAANSTLIQISGNVTADFIASA 807
QY 1372 --SDIVAHNLTLSGDIQTAKGASOANNSAGY-VDADGNKVIVYDSTDNKYQAKN-DGTYD 1427
DB 808 DGTGIVEFVNTGPIVNTLNKQAVPNALKOITVSGFNVVV-----NEIGNAGNYIGAMT 862
QY 1428 KTEKAVADKLVQAQATPD-----GTLAQMNVKSVINKEQVNDANKKQGINEDNAFVKG 1480
DB 863 DTIAFENSSLGAVLFLSPGPFNDAGNTIPLTIKSTVGNETA-----EGFSPSVIVSG 916
QY 1481 LEKAASDNKT---KNAAVTVG---DLNVAQTPITPAGDTGTAKKLGELTTLTKGGQDGT 1534
DB 917 VDSVIADQGVIGDQNNIVGLGSDNGIIVNATTLYAG-IGTINNQG--TTLTSGGVPNT 974
QY 1535 NKLDNNIGVVAGTDGP--TYKLAKDLNLNSVAGGTKIDE-----KGISFVDA 1582
DB 975 PG-TVYGLGTIGIGASKFKQVTFITDYNLGNLIATNTINDGVTVTVTGGIAAGGIAGTDF 1033
QY 1583 NGQ-----AKANTPVLSANGL-----DLGKVISNVGKCTKDDTAA 1618
DB 1034 DGKITLGSVNGNANVRFADGIFSNSTSMIVTTKANNGTFTVYLGNAFVGNI--GOSDTPVA 1091
QY 1619 NVQQLNEVRNLLGLGNDNADGNNOVNIADIKKDPNSSGSSNRTVIKAGTVLGK----- 1671
DB 1092 SVR-----FTGSNNGAGLKNIIYQVID---FGYVNLGIVNSVILGSGSTAINCK 1139
QY 1672 -----GNN-----DTEKLTG--GVQGVGDKDG-----N 1693
DB 1140 IDLLTNTLTFAGTSTWGNNTSIETTLTLANGNIGHIVIAEAGAOVNATTTGTTINVDN 1199
QY 1694 ANGDLISNVVVKTKDQSKK-----ALLATYNAAG 1722
DB 1200 ANANFSGTQTYTILGGAARFNGLGPNFTVTSNRFVNYGLIRAANDYVITRTNNAEN 1259
QY 1723 -QTNVYTNPAEAIADRINEQIRFFH-VNDGNQEPVQVGRNGIDSSAGKISVAIGFOAK 1780
DB 1260 IVTNDITNSPFGGAPGVQGVNTTFVNATNTAAAYNNLLAKNSADSANFVGTITVDTSAI 1319
QY 1781 ADGEAAVAGRTQAGNQSTAGDQNAQATGDSIAITGTGNVAGHSGAIGDPSVKA--- 1838

Thu Sep 13 14:19:31 2001

Db 804 TKNRGS-QTEVAGDLTITANK-----DLLEHGAHHVEGRYQESGENIQHLAVNDS-- 853
 Qy 1301 FANGNTTTAKVYDDTSKTK-----VYVDVNDVTIEVKDKKL---GVKTTTLTSTGTG 1353
 Db 854 -----ETSKTSLNIGDVGNDYSGVTKPKRAIEDGVNTT-----KP 893
 Qy 1354 ANKFAISNOATGDALVKASDIIVAHNLTL-----GDIOTAKGASQA---N 1395
 Db 894 GNNTDLTKKVT-----ARDIAIANLANLSNLETPNVGVEGKGGGQQSQTDSQAVSTS 947
 Qy 1396 NSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKVEAKDKLVAAQOTPGTL----- 1448
 Db 948 INAGKIDIDSNNKLHD--OGTHYQSTQEGI-----SLTANTHTSEATLDKHKQTFP 995
 Qy 1449 -----AQMNVKSVINKQVNDANKKOGINEDNAFVGLKKAASDNKTKNAAVTVGD-- 1499
 Db 996 HETKGGGIGVSTKGSQ--ITVALKGEQTTDNALMETKAKGSOFTSNGDISINVGNAH 1054
 Qy 1500 -----LNVAQAQPLTFAGPT-----GTTAKKLGTELTK-----G 1529
 Db 1055 YEGAQFDAGKQKTVINAGGDLTLAQATDTHSESQSNVNGSANLKVGTTPESKDYGGGFNA 1114
 Qy 1530 QOTDINK-LTDNNGVAGTGTGFTVKLAKDLT-----NLNSVNAAGTKID--- 1573
 Db 1115 GTTHHSKEQTTAKVGTITGSGQIEINAGHNLTLOGTHLSSEQDIALNATN---KVDLQS 1170
 Qy 1574 -----EKGISF---VDANGQAKANTPVLSANGDLGKVISNVGKTKOTDAANVQOLN 1624
 Db 1171 ASSEHTEKGNLGGVAGFGKMTDDASSVNL---GSAQFAIGKQDEKSVSRGGGTIN 1227
 Qy 1625 EVRNLLGLND-NADGNQVNIADIKKPNSG-----SSSNRTVKA--GTVLG--GKGNND 1675
 Db 1228 NSGNLTINGSVHLOGAQVNSKDTQLTSSQDIEITSQAQTDYKKNWGTIDIGFNCKTNN 1287
 Qy 1676 T-----EKLAT-----GQVQGVGD-----KGNANGDLSNVWVKTKD-----GSKKA 1713
 Db 1288 TPKEVTEERKPATSIHNIIGKLLVNDQOQKSHQATLETGLTINSNKDLTSLGANVTA 1347
 Qy 1714 LLATYNAAGQTVNTPNPAEADIRINEGIR--FFHVDNGOEPVVGVRNGRIDSSASG-- 1769
 Db 1348 DSVTGNVGGSLNTASQES---DRHVTGVNVGVNITNDPKSSQVKN-----TAKAGGSL 1399
 Qy 1770 -----KHSVAIGFQAKADGAEAAVAGROTOAGNOSTA-----TGDNAQATGDSIA-IGTG 1819
 Db 1400 LEKTIKTIDSGIKSTF-----AISDKYNSLSSTIADTKIGSDETKAKIDQGFQGVNG 1454
 Qy 1820 ---NVVAKHSGAIGDPSTVKADNSYSVGNNGNOFTDQTDVFCVGNNTVITNESNVALGS 1877
 Db 1455 IKNIVTG-----AEGHTANADIKVTHVDNDVTKTT-----SLTSNNDLSLNV 1497
 Qy 1878 NSAISAGTHAGTQAKSDGTAGTTTATAGATCTVKGFAGTAVGAVSVGASGAERRIQNVA 1937
 Db 1498 N-----GSEKLTGA--EIVSQOQGVDLGGSSV-----KLENIE 1528
 Qy 1938 ACEVSATSDAVNGSOLYKATQSTAN-----ATNE 1967
 Db 1529 GIIHYEAGADLDLKSVDLAKOLQGVGDLSFKSPVKTN 1566
 RESULT 6
 OMPB_RICKRI ID OMPB_RICRI STANDARD; PRF; 1654 AA.
 AC Q53047;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
 GN OMPB.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

Rickettsiaceae; Rickettsieae; Rickettsia.
 OC NCBI_TaxID=783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=92167802; PubMed=1724278;
 RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
 RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
 rickettsii is encoded by an unusually long open reading frame;
 RT evidence for protein processing from a large precursor.";
 RL Mol. Microbiol. 5:2361-2370(1991).
 RN [2]
 RP SEQUENCE OF 279-1654 FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=90136087; PubMed=2515418;
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
 RT "Cloning, expression and sequence analysis of the gene encoding the
 RT 120 kd surface-exposed protein of Rickettsia rickettsii.";
 RL Mol. Microbiol. 3:1579-1586(1989).
 CC FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X16353; CAA34403.1; --
 CC Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
 KW CHAIN 1 1333
 FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
 FT DOMAIN 1181 1188 POLY-THR.
 FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
 SQ
 Query Match 4.3%; Score 443.5; DB 1; Length 1654;
 Best Local Similarity 22.0%; Pred. No. 5.2e-09;
 Matches 383; Conservative 189; Mismatches 567; Indels 605; Gaps 92;
 Qy 650 SSGNITTPYNI-----SVKTTKLNSGTSGNNKFSVNAHDNNSLVTAADLYLNKY 703
 Db 13 SAGLVASTATIVASFAGSANGAAIQNRRT-----NGAATTVDGAGF----- 55
 Qy 704 NETADSLPSPKQVNGDNNNAITVVGKDTNCKTNT-----LKLKGENVNITNTRAT 756
 Db 56 DOTAAPA-----NVGVALNAVITANANNNGINENTPAGSFNGLLLTANNLAVTVSEDT 108
 Qy 757 --GTVTFGIDQSNGLTTPKLVGSDTNGNRLV1--EQVPSADGNSTKNITKGLSPTLPSI 812
 Db 109 TLGFTITNVVHAHSF---NLTLNA---CKTLTITCGQVTNAQAATKN----- 150
 Qy 813 ASPSGRNTALGNTIEEKDKSNAASIDDLVNLNAGFNKNGKDKDFVSTYDVFIDGNATT 872
 Db 151 ----AQNVVV-----QFNNGAAIDN-----NDLKGVGR-----IDFGAPA 181
 Qy 873 AIVTYDEANQTSKVAYDVNVDEKTIETLTDGNGKKOLGVK-TIKLTETSTNGNATTFSTD 931
 Db 182 STLVEFLNAPTT-----OKAPLIIGDNAVIANGVNGLVNTNGFIQVSNKSFAT-- 230
 Qy 932 DHALVKASDIAGNLNTLAEIHTTKGTANTALQOTFTVVKVVDENDKADDTNATVCKDCTS 991
 Db 231 ----VKAINLADQGLII---FNTDANNANTL-----NLQAGGTTINFTDGTG 272
 Qy 992 GKV-----NTLKLKGNGLDIK-----TDKDGTVT-----FGINTOSGLKA 1027
 Db 273 RLVLKSKHAAATNFNTIGSLGGLKGVIEFNTVAVDQGLTANAGAANAVIGTNNGAGRAA 332

QY 1028 GDSFTLNNGLSIKNTASNEQIOVGADGVKPA-----MVN-NGVVGAGIDGTT 1074
Db 333 GFWSDNGKAT-----IDGOVYAKDMVLOSANATGQVFRHIVDVGADGTT 380
QY 1075 RITRDEIGFTGTNGSLDKSPHLSKDGINAGKKTITQSGEIAKNSHDVATGGKIYDLK 1134
Db 381 AF-----KTAASKVITQDS-NFG-----NTDFGNLA-----AQ 408
QY 1135 TELENKISSTAKTAONSLHEFSVADEQNNFT--VSNPYSSYDTSKTSDDVTF 1185
Db 409 IKVPNAITLT-----GNETGDASNP-----GNTAGVITFDANGTLE 444
QY 1186 -----AGENGLITKYNKGVVRVIGIDOTKLTTPKLTGVGNNN-----GKGVINSQNG 1232
Db 445 SASADANVAVNTIITAIBASG---AGVVLQSGTHAAELRLGNAGSIFKLADGTVINGKVN 501
QY 1233 QNTITGLSNTLANVTNDKGSVRTTEQGN-----IIKDEKTRAASIVDVLSAG- 1280
Db 502 QPALVGGALAACTITLDGSATITGDIAGNAGGAALQRLITLANDAKKTLTLCGANIIGAGG 561
QY 1281 --FNLQNGEAVDFVSTYD--TVNFANGNTTAKVYDDTSKTSKVYDVNVDDTTIEVK 1336
Db 562 GTIDLOANGGTTIKLSTONNIWVDFDLAATDOTGWDASSLT-----NAQTLTINGK 614
QY 1337 DKKLGKVTTLTSTGTGANKFALSN-QATGDALVKASDIVAHLNLTSLGDIQTAKA---S 1392
Db 615 IGTIGANNKTLQGFNIGSKTSKVLNSGNVAINELVIGND-----GAVQFAHDTYLYIT 665
QY 1393 QANNSAGVVDAGNKVIVYDSTDKYQAKNDGTVDKTRVAKDKLVAQAOTPDGTFLAOMN 1452
Db 666 RTTNAAGQ-----GKLIEN-----PVVNGGTT-----LAAGTNLGSATNP--LAEIN 705
QY 1453 VKSVINKEQVNDANKKQGINEDNAP--VKGLEKAAKDNKTNAAV-----TVGDNLNAQA 1505
Db 706 FGS-----KGVNVDVLNVGEGVNLATNITTTDANVGSFVFNAGGTNIVSG 752
QY 1506 TP-----LTFAGDTGTTAKKLG-----ETLTIKGQDPTNKLTDNNIGV 1544
Db 753 TVGGQOGKNFTVALENGFTVFKFLGNATFNWNTTIAANSTLQIGGNYTADCVASADGTGI 812
QY 1545 V--AGTGDFTVKLADLTNLNSV-----NAG--GTKID----- 1573
Db 813 VEFVNTGPTITVNLKQAPVNAKQITVSGPGNVINELIGNAGNHGAVTDIAPENSLSL 872
QY 1574 -----EKGISFVDANG---QAKANTPVLSANGLDLGKVISNYGKGTOKTDDAANVOQL 1623
Db 873 GAVVELPRGIPENDAGNTMPLTIKSTVGNKTKAKGDFVPSVVLGV-----DSVIADQVVI 927
QY 1624 NEVRNLLGLGNDNAGQNVN-----IADIKKDPN-----SCSSSNR--TVIKAGTVLGG 1670
Db 928 GQNNIVGLGLGSDNGIIVNATLYAGISTLNNNGQTVTLGGVPTPGTVYGLGTGGA 987
QY 1671 K-----GN-----NDTEKATGQVGVGVDRKDG-----NANGD--- 1697
Db 988 SKFKQVFTTDDYNNLGNIIATNATINDGVTVTTGGI-AGIGDFGKITLGSVNGNGVREFA 1046
QY 1698 ----LSNVNWKQKSGSKALLATYNAAGQTNVYN-----NPAEAI-----DR 1737
Db 1047 DGILSN--STSMIGTKRANNGTVTVYLGNA-FVGNIGSDTDPVASVRFRTGSDGAGLQGN 1102
QY 1738 INEQGIRFHVNDG--NOEPVVQGR---NG-ID-----SSASKHSHVAIGFOAKAD--- 1782
Db 1103 IYSQVIDFGTVNLGIVNSNIIILGGGTTAINGKIDILVTNLTAFASGTSTWGNNTSTETTLT 1162
QY 1783 -----GEAAVAIRGQ-----TOAGNOSIAGDQNAQA--TGDOQSIAGTGNVYAKGHSAGT 1831
Db 1163 LANGNIGHIVILEGAQVNTTTTGTITIKVQDNANANFSGTQTYTLIQ-----GARENGTLG 1219
QY 1832 DP-----STVKADN--SYSVGNNNQFTDATQTDVF-----GYGNNTI--V 1867
Db 1220 SPNFAVTSNREVNYSLIRAAQNDVITRTNNAENVVINDIANSPFGGAPGVQNVITFV 1279
QY 1868 TESNSVA-----LGSNSALSA-----GTHAGTQAKK 1893

Db 1280 NATNTAAYNNLLAKNSANSANFVGAIVTDTSAATNVOLDLAKDIAQQLGNRLG--ALR 1337
QY 1894 SDGTAGTTTATGATGTVKGFAGQTAGVAGSVGASGAERIQNVAAAG-EVSATSTDA--- 1948
Db 1338 YLGTPTETAEMAG-----PEAGAISAAVAGDEALDNVAYGIWAKPFTYDAHQSK 1386
QY 1949 VNGSOLYKA-----TOSIAN-----ATNELDHRITHONENKANA-GISSAMA 1988
Db 1387 KGLAGYKAKTGTGTVIGLDTLANDNLMIAGAAIGITKIDIKHODYKKDKTVDVNGFSFLY 1446
QY 1989 MASMPQAYIPGRSMYTGTTATNGQCAVAVGLSKLSDNGQWYFKINGSADTQGHVGAAGV 2048
Db 1447 GA-----QOLVKNFFA-----QGSATFSLNOVKNKSORYF-----FDANGMSKQIA 1488
QY 2049 AGFH 2052
Db 1489 AGHY 1492
RESULT 7
P2P_LACLA
ID P2P_LACLA STANDARD; PRT; 1902 AA.
AC P15293;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE) (LP151).
GN PRT.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid pUp763.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCDO 763;
RC MEDLINE=89313288; PubMed=2501630;
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
from Streptococcus lactis NCDO763";
RL Mol. Microbiol. 3:359-369(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
INSULIN B-CHAIN.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X14130; CAA32350.1; -;
DR PIR; S06997; S06997.
DR HSP; P00782; 2SPT.
DR MEROPS; S08.019; -;
DR InterPro; IPR000209; -;
DR InterPro; IPR001899; -;
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.

Db	624	PFIAGSOALLQALNNKNNPFYAYYKQLKGALTJDLFLKTVEMNTAQPIINDIYNVVI--VS	682
Qy	683	NAHDNNSLVITAKDLADYLNKVNET--ADSALEPFKQVONGDNENNAITVCKDTNGKTFNPL	740
Db	683	PRGAGLVDAKAAIDALEKNPSTVVAENGYPAVELKDLFTSD-----KTF-----	728
Qy	741	KLKGENGVNITNKRATGVTVTFGIDQSNGLTTPKLTIVGSDTNGRLVIEOVPSADGNSTKN	800
Db	729	KL-----TFTNRTHLTYQMD-SNTDTPNAVYT SATDPNSCVLYDKKIDGA-----	773
Qy	801	LIKGLSPTLSIASPSGRNTALGNTIE-EKDKSNAASIDDLVN-----AGFNLK-----	848
Db	774	AIKAGS-----NITVPACKTAQIEFTLSLPKSFQDQOQFVEGLFNKSGDSRLNLPYMGFF	829
Qy	849	-----NNGKDKFVS-----TYDVTDFIDGNATPAT-----VNYDEANOT---SKVA	887
Db	830	GDWMDGKIVDSLNGITYSPAGNGFTVPLLT-NKNTQTYQYGWTVTDAGDGNQIVDDQATA	888
Qy	888	Y-----DV-----NVDEKTTELTGDNKKOLGVKTKLKTETSTNGNATTFSTD	930
Db	889	FSSDKNALYNDISMKYLLLRNISNVQVDILDGQGNK-----VTTLSSTNRRKTTYNAH	942
Qy	931	DPHAL-----VKASDIAGNLUNTLAEIHITTKGTANTALQFTTVKKVDE--NDKAD-----	978
Db	943	SOQYIYHAPAWD-----GTYYDQRDGNIKTADGGSYTYRISGVPEGDKRQVDFPFKL	997
Qy	979	DPNATLYVGKDGTSKGVNPLKLGKNGLDIKTKDKQ-----TVTFEINTQSGKA--GDST	1031
Db	998	DSKAPTIVRHVALSAKTENGKTYLYLTAEKADDLGLDGLATKSVKTAINEVTLNLDATFTDAG	1057
Qy	1032	TLNNGLSLKNTASNEQTOVGADGVKFAWNVGVGAGIDGCTTRITRDEIGFTCTGSLD	1091
Db	1058	TTADGTYKLETLSLSDQAQ-----LGNG-----DMSAELYLDNASNATDQDAS	1102
Qy	1092	KSKPHLSKDG--INAGG--KKITNQSGEIAKNSHDAVT-GCKYIDLKTELENIKLSSTAK	1146
Db	1103	VQPGSTFSDLVINGGGIPDKISSITTG-----YEANTQGGVTFESGTYPAAVDGTYT	1156
Qy	1147	TAQNSLHESVA--DEOGNFTVSNPYSSVDTSKTSDVITFACENGI---TTKVNKGVVVR	1201
Db	1157	DAQCKRHDLNTTYDAATNSFTASMPVTNADYAAQVDLTADKHAHTQLKHFDTKVR-----	1211
Qy	1202	VGIDQTKGLTTPKLT--VGNNNKG-----GIV-----INSQNGQNTITGL-----	1239
Db	1212	-----LTAPTFTDLKENGSDQTSSEATIKVTGTVSADTFTVNVGDTVAALDAQHHFS	1263
Qy	1240	-----SNTLANVNDKGSVPTTEQGNLIKDEDKTRAASIVDVLSAGFNLQGNCEAVD	1291
Db	1264	VDVPVNTGDNTEIKVATFDEGNTTITQKTTSSYDP-----DMLK-----NSVTFD	1309
Qy	1292	FVSTYDTPVNEFANGNTTAKVTYDDTSKTSKVYVDVNVDDTTEVDDKKLGVC-----	1346
Db	1310	QGVTFGANEF---NATSAKE-YDPKTCIAITITKGVKHPHTTTLQVDDCKQIPKDDLTFSFT	1365
Qy	1347	LTSTGTCANKFAL-----SNOATGDALVKASDIVAHLNLTLSGDIOTAKGASQANNAGY	1400
Db	1366	LDLGLTLGCKPFGVVGVDDTQNKTFQEALTFILDAVA--PTLSLDSST-----DAPVY	1415
Qy	1401	VDADGNKVIYDSTDN-KYYQAKNDGT-----VDKTEKVAKDKLVIAQAQTPDGTLAQMVV	1453
Db	1416	TNNPFIQTGATDNAQYLSLSINGSSVSAQYVDINTNSGKPGHMAIDQ-PVKLLBEGKNV	1474
Qy	1454	KSVINKEOVDANKKQGINEDNAFVKGLEKAASDNKTKNAAVTVGDLNAVAQPTPLTFAGD	1513
Db	1475	LTV-----AVTD-----SEDNTTKNITVYEPKTKL-AAPTIV-----TP-----	1508
Qy	1514	TGTTAKKLGETLITKGGOTDNKLTNNIGNVVGATDGTFTVLAKDLTLNLSNVNAGG--T	1570
Db	1509	STTEPAKTVTLTANSAAT-----GETVOYSAD-----GGKTYQ	1541
Qy	1571	KIDEGISFVDANGAOKANTPVLNGLDGLGKVISNVGKTKDTDAANVQOL-NEVRNL	1629

QY 1630 LGLGNDNADGNQVNIADIKDPNSGSSNRRTVIKAGTVLGGKGNDEKLTATG--GVQVG 1687
 Db 1597 -----IASAKTSLASGKYDDATT-----TALAAA-----TQKAQTALDQTNAS 1634
 QY 1688 VDKDGNANGDLNVMVYKTKDGSKALLATYNAAGTQNVVTNNPAEADIRNEQGIREFH 1747
 Db 1635 VDSLTCANRDIQT-----AINQLAKLPADKKTSL--NQLQSVKAALETDL---- 1679
 QY 1748 VNDGNOEPVVOGRNGIDSSASGKSHVAIFGQAKADEAAVAIGROTQAGNQSIAIGDNAQ 1807
 Db 1680 ---GNO-----TDSSTGK-----TFTAALDDLVA-----QAQACTQT---DDQLQ 1713
 QY 1808 ATGDQSIAIGTGNVAGKHSGAIGDPSTVKADNSVSGNNOFTDATQDVFVGVGNITIV 1867
 Db 1714 AT-----LAKVLDAVLAKLAEG-----IKAATPAEVGNK-----DAATGKTWYADIADTL 1759
 QY 1868 TESNSVALGSN-----SAISAGTHAGTQAKSDGTAGTTTTAGATGTVKGFAG 1915
 Db 1760 TSGQASADASDKLAHLQALQSLTKVAARAEAKTVVGKDGTTGTGSDKGGGQGTAPTPTG 1819
 QY 1916 OPVAGVSGVSGASGAEIRIQNVAAVEV-----SATSTD 1947
 Db 1820 -----DIGKDKGDEGSPSSGGNIPTNPATTITSTSTD 1851
 RESULT 8
 ID OMPB_RICTY STANDARD; PRT; 1645 AA.
 AC P96989;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
 GN SLP.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 ON NCBI_TaxID=785;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WILMINGTON;
 RX MEDLINE=94040787; PubMed=8224886;
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
 RT "Cloning and sequence analysis of the gene encoding the crystalline
 RT surface layer protein of Rickettsia typhi";
 RL Gene 133:129-133(1993).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN=WILMINGTON;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNR fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii";
 RL Mol. Immunol. 29:95-105(1992).
 RN [3]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent
 RT mutant deficient in processing";
 RL Infect. Immun. 60:159-165(1992).
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN AND/OR MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; L04661; AAB48987.1; -
 KW Antigen; S-layer; Transmembrane; Cell wall.
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;
 Query Match 4.2%; Score 436.5; DB 1; Length 1645;
 Best Local Similarity 21.8%; Pred. No. 9.2e-09;
 Matches 366; Conservative 185; Mismatches 566; Indels 563; Gaps 83;
 QY 650 SSGNITPTPTNISV-----KTKLNSGTSGNNKFSVSNADHN-----NSL 690
 Db 13 SAGLVASTATIVAGFSGVAMGVMQVNRNTTNAATVVDGAG-FDQTGAGVNLPAVNSV 71
 QY 691 VTAADLADLYLNKVNETHADSALPSFKVQNGDNSNNAITVGKDTNGKTFNTLKLKGENVNI 750
 Db 72 ITA-----NSNNAITF--NTPNGNLNSLFLDTANTLAV 102
 QY 751 TYNRAT--GTVPFGIDQSNGLTTPKLTVGSDTN---GNRLVIEQVPSADGNTKNIKG 804
 Db 103 TINENTTLGFVNTVKQGNFF---NETIGAKSLTITGHGITAQQ--AATKSAQNVS- 156
 QY 805 LSPTLSIASPSGRNIALGNTIEEKDKSNAASID----DVLNAGFNLN----- 849
 Db 157 -----KVNAGAAIINDNDLSGVGSIDFTAAPSVLE--FNLIPTTQEAPLTL 200
 QY 850 -----NG-----KDKDFVSTYDTVFIDGNATTAIVTYDEANQTSKVAY 888
 Db 201 GDNAKIVGANGILNITNGFVKVSDKTFAG-IKTIIGNDQGLMFTNTPDAAN-----AL 254
 QY 889 DVNVDEKTIELTDGNGKQOLGVKTIKLTETSTNGNATTFSTDDHALVKASDIAGNLNTL 948
 Db 255 NLOGGNTINFNGRDG-----TGKLVLSKNGNATEFNV-----TGS LGGNLKV 299
 QY 949 AE-----EIHTKTGTANTALQT-----FTVKKYVDENDKADDTNAITVGKDGTS 991
 Db 300 IEFDTTAAAGKLIANGGAANAVIGTDNGAGRAAGFIV-----SVDNGNAATI-----S 347
 QY 992 GKVTILKLGKNGLDIKTDKDGTVTF-----GINTQSLKAGDSTLNNGLSIKNTAS 1045
 Db 348 GQV-----YAKDIIQIOSANAGGQVTFEHLVDVLGGKTNFKTADSKVI-----ITENAS 396
 QY 1046 NEQIQVAGDGVKFAVMYNGVVGAGIDGTTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAG 1105
 Db 397 FGSTDFGNLAVQIVVFNKI-----LTCNFTGDAKNNGN----- 430
 QY 1106 GKKITIQSGEIAKNSHDVATGGKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNF 1165
 Db 431 -----TAGVITFNANGTLVSGN-----NIVVTNIKATEVEGAGIVQLSGHGAELRLGAGSIF 447
 QY 1166 TVSNPYSSDTSKTSVDITFAGENGITTKVKN--GVYRVGIDQTKGLTTPKLTGVNHN--- 1221
 Db 448 --TDP-----NIVVTNIKATEVEGAGIVQLSGHGAELRLGAGSIF 487
 QY 1222 --GKIVINSQGNITITGLSNTLA--NVINDKGSVTRTBOGN-----I IKDEDKTRAAS 1272
 Db 488 KLDGTVINGPVNQNLPLVN--NNALAAGSIQDLSAIIITGIGNGAVNALQDITLANDAS 546

FT REPEAT 1732 1751 2-15.
FT REPEAT 1753 1772 2-16.
FT REPEAT 1795 1814 2-17.
FT REPEAT 1820 1839 2-18.
FT REPEAT 1840 1859 2-19.
FT REPEAT 1861 1880 2-20.
FT REPEAT 1887 1906 2-21.
FT REPEAT 1908 1927 2-22.
FT REPEAT 1929 1948 2-23.
FT REPEAT 1969 1982 2-24 (APPROXIMATE).
FT REPEAT 1983 2002 2-25.
FT REPEAT 2008 2027 2-26.
FT REPEAT 2028 2047 2-27.
FT REPEAT 2051 2070 2-28.
FT REPEAT 2071 2090 2-29.
FT REPEAT 2093 2112 2-30.
FT REPEAT 2120 2139 2-31.
SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;

Query Match 4.2%; Score 432.5; DB 1; Length 2334;
Best Local Similarity 20.3%; Pred. No. 1.9e-08;
Matches 466; Conservative 272; Mismatches 781; Indels 779; Gaps 118;

QY 182 LKXIQTSTGKIKYRTRAGCHAST-AVGA-MSYAOGHFSNAECTATAPAAYSLAVGLA 239
DB 369 LKPIQNTVTKATLKTYVAHSYGTGTGLWLDTVNSYDNKATVNTWTKPASKNIG--- 424

QY 240 AQATKSSIAVGSNAKANAFATAIGNTVNLG-----RGVALFGSOTLDRDNT--- 291
DB 425 -----RADVHKGQWASVDYTAAVKSWNSGGANYGFKLHTNGNGKEYWKLISSANSANKP 479

QY 292 --DASAYVPLGKTLADQYKATROGSDTDFISGNSNNSSIRKLIINVAGSRDRTDAVN 349
DB 480 YIEVTVTIPIKGN--PTIRAYHNGDSTGYDFISWKKEVGAQYKVIWYNGKEYQOALISAGN 537

QY 350 VAQIK-----LVBELANRKITFFKGDGDNNSNSVERGLGNTLTITKGDQATNALTEAN- 400
DB 538 VTSWSTGKKIWPSTAETASIAKRYKLHLGD---KDGAEALDPSVPYKNSGGSGYA-TSKNY 593

QY 401 -IGV---VTDNGKLKVKLAKELTGLTSVSATNKITVSNNTNNNAELQSG--GLTFSPITG 454
DB 594 WIGVSAIFDQEGEAMSAKAPV--IPNVGKAQAPSAKGYNNGNA---TGVFDLSWRAVSG 648

QY 455 TKTDKTVYSIDGLKF---TNDNSNIATKGTTRITRKKKIGFAGTNGDVDESKPYLDNEKL 510
DB 649 A-TGYKVQVFNKGFFETLDLGNQTSWTTKG-----KKI-----WPTSAEI 687

QY 511 KVCNSTLN-----SGS-LTVNNVTGNKQIQVGANGIK-----FATVANN-----VANTSATV 556
DB 688 KAGYALHLKDGSGAELPIN--PGPTYKNAGDGAQRNYSFKIAYNKGGEAIASPAATP 745

QY 557 GTARITEEK--IGFAGTNDGVDEQAPYLDKERLKVGRVEITTDGGINAGNHKITGLTNGI 614
DB 746 ALPDIARPKNVWTGILYTN-----TKSSQGTGVNLIWEKVQNAKGYK-NIYNG- 792

QY 615 ANTDAVITKQKDAKPLNAGDGISNSNNGDLVDSGNITPTTYNISVYKTKLNSNGTS 674
DB 793 -----KEYQ-----SFDVGDADHWTQNKNIWPTSEEIKAGSY-----KLHTDGGK 833

QY 675 GNNKFSVNAHDNNSLVATKDLADLYLNKVNETAADALPSFKVONGDNNNAITVKGDKTNG 734
DB 834 GELALDPSVYNNAN-----GNRYGKNY-----SFTL-----VAYDANG 868

QY 735 KTFNTLKLK---GENGVNITFNATGTVTFTGIDQSNGLTTPKLTGVSDTNGRNLVIEQVP 791
DB 869 ETIPTAPFNPTFHEGAFFLCTEYEWLSIIDIPSCQLNGAT-----GNVIVNEEDL 917

QY 792 SADGNSTKNIKGLSPTLPSIAS----- 814
DB 918 SIDGRPG---LGLSRTYNSLSDSDHLFGQGWYADAETSVISTDGGAMYIDEDATTHFT 974

QY 815 -----PSGRNIALGNTIEE---KDSNAASIDDLVNLNAGFNLNKNGKDFVSTYDT 862

DB 975 KKADGTYQPTGVILELTETADQFILTK-----DOTNAYFNKKGGKLOK----- 1019

QY 863 VDFIDGNATTATVTYDEANOTS-----KVAYDVNVNDEKTIETLTDGNGKKQOLGVYTIK 914

DB 1020 --VVDGHNNATVTYNDKNQLTAITDASGRKLTFTYDENGCHVTSITGPKNKKVYTSYEND 1077

QY 915 LYE--TSTNGNATFTSDDHALVKASDIAGNLNTLAEEI-----HTTKGTANTALQ 964

DB 1078 LLKKVTDGTGTVTSYDYDSEGRLVK--QYSAN--STEAKPVFTEYQYSGHRLERAINAKKE 1134

QY 965 TFFVKKVDENDKA-----DDTNAITVKGDKTSGKVNTP--LKLKKNGL-- 1005

DB 1135 TY-VYSYDADKKTLTLLMTQPNGRKRVQYGYNEAGNPIQVIDDAEGLKIKTNTKYEGNNVED 1193

QY 1006 ----DKT-----DKDGTVT-----FGINTQSLKAGDSTLLNNGLSIKNTA--- 1044

DB 1194 VDPNDVGTGKATESYQYDKDGNVTSVKDAYGTETIYFNKNDVTKMKDTEGNTVDIAYDG 1253

QY 1045 ----SNEQIOVGADGVKFAVNNNGVGA-GIDGTTTRIRD---EIGFTGTNGSLDKSKP 1095

DB 1254 LDVSETDQSGKSSAAVYDKYGNQIOSSKDLASTNILKDGSGFEAQKSGWNLTASKDRR 1313

QY 1096 HLS----KGINAGGKI-----TNIOSGEI----AK 1119

DB 1314 KISVIADKSGVLGSKALEVLSQSTSAGTDHGYSSATQTVLEPNTTYTILSGRIKTDLAK 1373

QY 1120 -----NSHDVGT----- 1127

DB 1374 SRAYFNIDLRDQKRIQWIHNEYSALAGKNDWTKROIPTTTPANAGKAVVMEVDHDK 1433

QY 1128 ---GKIYDLKTELE-NKISSTAKTAQNSLHEFSVADQGNNTVSNPYSSYDTSKTSDDVI 1183

DB 1434 DGKGAWFDEQLEKEGVSSSPNVQNS--SFTSATE---MNVSG--ASVDS----- 1479

QY 1184 TPAECNCITTKVKNGVVRVIGIDQTKGLTTPKLTGVGN--NGKGIVINSQNGQNTITGLS-- 1240

DB 1480 ---EEGFNDVSLKAARTSASQAGSVTKOTVVLGOSANDKPVYL-----TLTGMSKA 1528

QY 1241 -----NTLANVTNDKGSV-----RTTQGNIIKDEKTRAASIVDVLSAG 1280

DB 1529 SSVKFTDEKDYSLQANVTYADGSTGIYNAKFPSCQEWNRANAAVPIKTPINKVDI--SIL 1587

QY 1281 FNILOGN-----EAVDFV-----STYDVTNFANGNTT-----AKVYDDT--SKT 1319

DB 1588 FOKSATGTVWFDDIRLIEGSLITKSTYD---SNGNVVTRKEDEBELGYATSTDYDETCKKT 1643

QY 1320 SKVVYDVNVDDTII-----EVKDKKLGCVKTTTLTGTGANKFALSNOATGD----- 1366

DB 1644 SET--DAKGEKTTVYTDADQDLTNMTLSNGTSLHSDYKDEGNEVSKTIRAGADQTYKFEY 1701

QY 1367 ---ALVKASDIVAHLNLTLSGDIQTAKGASANNAGSYVDADGNKV--IYDSTD---NKY 1417

DB 1702 DVWGKLKTTTDPILG--NVLASEYDA-----NSMLTTTISPNGNEVSLSDGTDTRVSKS 1753

QY 1418 YQA--KNDGTVDTKTEVAKDKLVAQAQTPDGLAQMNVKSVINKEQVNDANKKOGINEDN 1475

DB 1754 YNGTEKYIFTYDKNG-----NETSVNKEQ--NTTKKRTEDKN 1790

QY 1476 AFVKGLEKAASDNKTKNAAVTVGDLNVAOPLTFAGDTGTTAKKLGLETLIKGGQDFTN 1535

DB 1791 RLTELTDRGGS-----QT-WTYPSP-----SDKLKTFSIHGDQGTN 1827

QY 1536 KLTDDNNGVYVAGTDGFTVKLAKDLTLNLSVNAAGTKIDKEG--ISFVDANGQAOKANTPVL 1593

DB 1828 QFTYN-----KLDQMIEMKDDSTSSYFDDENGQVQFTITGNG----- 1865

QY 1594 SANGDLGGKVINVGKGTKDTDAANYOQLNEVRNL---LGLGNDNADGNQVNTADIKD 1650

DB 1866 -----GGTSFS-----YDERNLVSSHLHIGDKN--GGDILTESYEYD 1899

QY 1651 PMSGSSSNRPVIK---ACTVL--GGKGNNDTEKLTATGCVQGVGDKGNANGDLSNVWKT 1705

Thu Sep 13 14:19:31 2001

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M24767; AA11677.1; -
CC HSP: Q99405; IMPT.
CC MEROPS: S08.019; -
CC InterPro; IPR000209; -
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00882; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 4.1%; Score 423.5; DB 1; Length 1902;
Best Local Similarity 20.7%; Pred. No. 3.1e-08;
Matches 457; Conservative 235; Mismatches 851; Indels 663; Gaps 99;

QY 20 AYAASHSTGGSCATGQVSVRTLSFARIAALAVLVIGATLNGSAYAOQITTK-IEIGQ 78
DB 31 AKAAISQQTGKSSLA-----NIVTAATAKAQATDTTAATTN-QAATATLAARKGIDYK 82
QY 79 TNKI-----NNTLKDALATGASTAFGSLSKNAGQOATAIGSVK 118
DB 83 LNKVQOQDIYVDVIVMSAAPASENGTLRTDYSSTAEIQETNKVIAQASVKAARVQVT 142
QY 119 PDPNNGSNGV-----GSHAKGNESIAI----- 141
DB 143 QQTAGESYGVVNGFSTKVRVVDIPKLAQIAGVKVTVLAKVYPTDAKANSMAVQAVWS 202
QY 142 -----GGDLAEGDASTAGSDDLILPKNLDLKNFHLKIHGHEILKIKIOTSTDG----- 191
DB 203 NYKKGEGTVSVSDISDIPTHKMDRLSDDKDKV-----LTKSDVEKFTDTAKHGRYFN 256
QY 192 -KIKYRTTAAOGHASTAVGAMSYAOG-HFSNAFGTYATAEAYSLAVGLAAQA----- 242
DB 257 SKVPYGFNTADNDITDDTVDEQHGIMHVAGTIGANGTDDPAKSVGVGAPEAQLAMKV 316
QY 243 --TKOSSIAVGSNAKANAFATAIGNTVVNLGRGVAGLFCGSQLDRDNNNTDASAYVPLG 300
DB 317 FTNSDFTSATTGSSLTLSAIEDSAKIGADVLNLSLG--SDSGNQTLE-----D 361
QY 301 KTLADYKATROGSDTDFISGNNSNNSSIRRKINVGAGSRDTDVAVNQLKLEELA 360
DB 362 PELAAVONANESGTAA-VISAGNSGTSGSA-----TEGVNK----- 396
QY 361 NRKITEFGDGDNNNSVERGLGNTLFIKGAQNTALTEANIGVVDGNGCLKVYLAKELTG 420
DB 397 ----DIYGLQDNEWVGTPGTSRGATTVASAENTDVITQA--VTITDGTGLQ----- 442
QY 421 LTSVSATNKITVNTNNAELQSGGLTFSPITGKTDTKTVSIDGLKFTNDSNSTIA--- 477
DB 443 -----GPGTIQLSSND-----FTGSDOKKFFVVKDASGNLSKGLADYVT 482

DB 1900 AN-----GNRTTINSASGKVOYEGKLNQLKETHEDGTVEITYDGFGN-----RKVTVT 1951
QY 1706 QKDSKALLATYNAAGQTVNTNPAEADRINEOGI-----RFFHVND--- 1750
DB 1952 IKDSSKTVNASFIMNQ-----LTKVNDESISYDKNGNRTSDGKFTYTDAED 2000
QY 1751 -----GNOEPVV-----QGRNGIDSSASGK-----HSVAIGFOAKADGEAAVAI 1789
DB 2001 NLTAVTKGGEOKPATYKYDEKGNRIQKTVNGKVTNYFYDGDLSNLVLYETDADNV----- 2056
QY 1790 GRQTOAGNQSIGDUNAO-----ATGD-QSIAIGTGNVAGKHSAGID 1832
DB 2057 -----TKSYTYDGSGLLSYENGKKYFYHNAHGDIIDISDSTGKTAKYQYDAWGN 2109
QY 1833 PSTVKA-----DNSYSVGNNOFTDAT-----QTDVF-----GVGNNTVTES 1870
DB 2110 PKTASDEVKDNRYA-GYQYDEETGLYLMARYEPRNGVFLSLDPPDPSGDSDLQ 2168
QY 1871 NSVALGSNS-----AISAG--THAGTOAKKSDGTAGTTTAGATGTVKGEA 1914
DB 2169 NGYAVGNPNVNVDPDGHVWLVVYNAGFAAYDGYKAYKSG-----KGWK 2213
QY 1915 QOTAVGAVSGASGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQSIANATNELDHRHQ 1974
DB 2214 GAWAAA-----SNEFGPKIKF-----GASRAYKFTKKAVKITGHTRHGLNQ 2255
QY 1975 N-----ENKANAGISSAMAMASMPQA---YIPGRSMVTGCIATHNCOGAVAVGL 2020
DB 2256 SIGRNGRGVNLRAKUNA-VRSPPKVIQPGATKYVCKKATV-----VLNKRKGVITATYS 2311
QY 2021 SKLSDNGQWFKINGSAD 2038
DB 2312 SR-AGSKHVFHTHGKN 2328
RESULT 10
ID_PIP_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE
DE PROTEINASE).
CN PRTP
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pWV05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WG2;
RX MEDLINE=88149035; PubMed=3278687;
RT Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris w92.";
RL Appl. Environ. Microbiol. 54:231-238 (1988).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

QY 478 --TKGTRITKK-KIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGLTVNNTGNKQIQ 534
 Db 483 ADAGKIAIVKRGELSP-----DDKOKYAAAGAAGLIIVNN----- 519
 QY 535 VGANGIKFATVANNVANTSATVGTARITBEKI-----GFAGTN 572
 Db 520 ---DGTATPVTSMALITTPFTFGLSSVTGQKLVDMWTAHPDDSLGVKIALTLVPNOKYTE 576
 QY 573 DGVEQAPYLDKRLKVRVEITDGSINAGNHKITGLNGIANTDAVTKIKOLKDAKPTL 632
 Db 577 DKMSDFTSY-----GPVSNLSFKPITAPGNGIWTQNNNGYTNMSGTSM-----ASPEI 626
 QY 633 NAGDGI---SINSGNDLVDSSNIITPTYNISVKTKLNS-----NCTSGNNKFSVSNH 685
 Db 627 AGSQALLKQALNNKPNFYAYYKQLKGTALTDLFKTVMNTAQPINDINYNVI-VSPRR 685
 QY 686 DNNSLVTAADLADYLNKVNET---ADSLPSFKVQNGDNSNNAITVGKDTNGKTFNTLKLK 743
 Db 686 QGAGLVDDVAAIDALEKNPSTVVAENGYPAVELKDFSTSD-----KTF---KL- 730
 QY 744 GENGVNITNRTAGTVTFGIDQSNGLTTPKLVGSDTNGRLVIEQVPSADGNSTNIK 803
 Db 731 ---TFTNSTTHELTQMD-SMTDTNAVYTSATDPNSGLVYDKKIDGA-----AIK 776
 QY 804 GLSPTLPSTASPSGRNALGNTIE-EKDSKNAASIDDVLN-----AGFNLK----- 848
 Db 777 AGS-----NITVPAGKTAQIEFTLSLPSFKDQOQFVEGFLNFKSGDSRLNLPYMGFFGDW 832
 QY 849 NNGKDKDFVS-----TYDVFDFIDGNATTAT-----VTYDEANOTSKVAYDVNVD 893
 Db 833 NDKIVDSLNGITYSPAGNFGTVPLLT-NKNTGTOYGGWVTDAGNQI-----VD 883
 QY 894 EKIETLTDGNGK-KOLGVKTIKLTSTNGNATTFSTDDHALVKASDIAGN-LNTLAE 950
 Db 884 DOAIAFSSDKNALYNDISMKYLLRNIS-----NVQVDILQOGKNTVTLSS 930
 QY 951 EIH7TK-----GTANTALQTFVKKVDE-NDKAD- 978
 Db 931 STNLTKTYNNAHSQOYIYNAPAWDGTYYDQRDGNKLTADDGSYTYRISGVEGGDKQV 990
 QY 979 -----DTNAITVKGDTSGKVTNLKLGKNGLDIKTKDG-----TVTFGINTOSGLK 1026
 Db 991 FDVPFKLDSKAPVTHVALSAKTENGKTOYLLTAEKADDLGLDASKVKTAINEVNLD 1050
 QY 1027 A--GDSTTLNNGSLTKNTASNEQIOVGAGVKFAMVNGNVGAGIDGTTTRITRDEIGFT 1084
 Db 1051 ATFTDAGTTADGYTKIETPLSDEQAQ-----LGNG-----DNSAELYLTDNASN 1095
 QY 1085 GTNGLDKSKPHLSKDG--INAGG--KKITNIOGSEIATKNSHDAVT-GKIIYDLKTELEN 1139
 Db 1096 ATQDDASVQKPGSTFSLIVNGGIPDKISSTTTG-----YEANTQGGGTFTFSGTYP 1149
 QY 1140 KISSTAKTQNSLHFSVA-DEQGNFTVSNPYSYSDTKTSKDVITFAGENGIT---TTK 1194
 Db 1150 AVDGTYNAGKKHDLNTTYDAATNSFTASMPVTNADYAAQVDLYADKAHTQLLKHFDTK 1209
 QY 1195 VNGVVRVIGIDQTKGLTTPKLT-VGNNGNK-----GIV-----INSQNGONTITGL 1239
 Db 1210 VR-----LTAPFTDLTKFNNGSDQTSATIKVTGTVSADTKTVNVGDTVAAL 1256
 QY 1240 -----SWLANVTNDKSVRTTEOGNIKDEKTRASIVDVLASAGFNQ 1284
 Db 1257 DAQHHSFSDVPVNVGNTIKVATDEGNTTTEQKTISSYDP-----DMLK----- 1303
 QY 1285 GNGEAVDFVSTYTVNFANGNTTAKVYDDTTSKTKVYDVNVDDTTIEVKDKLGVK- 1343
 Db 1304 -NSVTFDOGVTFGANEF---NATSAKF-YDPKGTGIATITGKVKHPTTLQVQDKQPIKD 1358
 QY 1344 -----TTTLTSTGTGANKFAL-----SNQATGDALVKASDIVAHLNLTSGDIQTAKGASQ 1393
 Db 1359 DLTFSTLGLTGLQKPFVGVGDDTTQNTKFQEALTFILDAVA--PTLSLESST----- 1410
 QY 1394 ANNSAGYVDADGNKVIYDSTDN-KYYQAKNDGT-----VDKTYEAKDKLVAAQATPDG 1446

Db 1411 --DAPVYNDPNFOITGTATDNDAQYLSLSINGSSVASQYVDININSCKPGHMAIDQ-PVK 1467
 QY 1447 TLAQMNKSVINKEQVNDANKKOCINEDNAFVKGLEKAAASDNKTNAAVTVGDLNAVAQT 1506
 Db 1468 LLEGKNVLTV-----AVTD-----SEDNTTKNITVYEPKKTAL-AAPTIV-----T 1507
 QY 1507 PLTFAGDTGTTAKKLGTELPIKGGQDTNTKLTNNIGVAGTGTFTVKLAKDLNLSVN 1566
 Db 1508 P-----STTEPAKTVTLTANSAAT-----GETVOYSAD----- 1535
 QY 1567 AGG---TKIDEKGISFVDANGQAKANTPVLSANGCLDGGKVISNVGKGTGTDTAANVQOL 1623
 Db 1536 -GKTYODVPAAGVT-VTANGTFKFKSTDLGYNESPAVDYVVTNI-----KADDPAQLOAA 1589
 QY 1624 -NEVRNLLGLGNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKGNNDPEKLTATG 1682
 Db 1590 KQELTNL-----IASAKTSLASGKYDDATT---TALAAA---TQKAQTA 1627
 QY 1683 --GVQGVVDKGNANGDLNWNVKTQKDGSKALLATYNAAGQTNVYNNPAAIDRINE 1740
 Db 1628 LDOTNASVDSLITGANRDLQT-----AINQLAAKLPADKKTSLL----- 1665
 QY 1741 QGIRFFHVNDGQNPVVOGRNGIDSSASGKHSVAIGFQAKADGAAVAIGRQTQAGNOSI 1800
 Db 1666 -----NOLQSVKAALGTDLGNQDTPSTGKTFTAAALDDLVA-----QAQAGTQT- 1708
 QY 1801 AIGDNAQTGDQSTAIAGTGNVAGKHSAGIDPSTVKADNSYSVGNNNQFTDATQTDVFG 1860
 Db 1709 --DDHQAT---LAKVLDVAVLAKLAEG-----IKAATPAEVDGNK---DAATGKTWY 1752
 QY 1861 VGNNTVTSNSVALGNS-----SAISAGTHAGTQAKKSDGTAGTATTTAGATG 1908
 Db 1753 ADIADTLTSGOASADADSKLAHLQALQSLKTKVAAAVEAAKTVGKGDTTSDKGGGQ 1812
 QY 1909 TVKGFAGQTAGVAVSVCASGAERRIQNVAAGEV-----SATSTD 1947
 Db 1813 TPAPAPG-----DIGKDKGEGSQSPSGGNIPNTNPATTTSTSTD 1851
 RESULT 11
 P3P_LACLC
 ID P3P_LACLC STANDARD; PRT: 1902 AA.
 AC P15292;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
 DE ASSOCIATED SERINE PROTEINASE).
 GN PRTP.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OG Plasmid.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus
 OC NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
 RC STRAIN=SK11;
 RC MEDLINE=89340435; PubMed=2760036;
 RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
 RT "Primary structure and organization of the gene for a procaryotic,
 RT cell envelope-located serine proteinase";
 RL J. Biol. Chem. 264:13579-13585(1989).
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; J04962; AAA03533.1; ALT_SEQ.
DR PIR; A32634; A32634.
DR HSP; P00782; 2SPT.
DR MEROPS; S08.019; .
DR InterPro; IPR000209; .
DR InterPro; IPR001899; .
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00882; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 4.1%; Score 423.5; DB 1; Length 1902;
Best Local Similarity 20.4%; Pred. No. 3.le-08;
Matches 458; Conservative 230; Mismatches 817; Indels 735; Gaps 104;

QY 20 AYA KSHSTGGSCATQGVSRVTLSPFARIALAVLIGATLNGSAYAAQOITTK-IEIQQ 78
DB 31 AKA AISQOTKGSSIA-----NTVTAATAKQAATDTTAAATN-QAIALQAAGIDYK 82
QY 79 TNKI-----NNTLKGDALEAGSTAFSLSKAQSQAIAIGSVK 118
DB 83 LNKVQOQDIYVDIVQMSAAPASENGILRTDYSTAEIQOETNKVIAAQAQVAAEQVT 142
QY 119 PDPNNGSGNV-----GSHAKGNESTAI----- 141
DB 143 QQTAGESYGVVNGFSTKVRVVDIPKLQIAGYKTVTLAKVYPTDAKANSMAVQAVWS 202
QY 142 -----GGDLVLAEGDASTAIGSDLLYLPKNLDLK-----NEFKLI-HGHEILKKI--- 185
DB 203 NYKYKGEVTVSVIDSGIDPDKHMLRSDDKDKVLRKSDVEKFTDVKHGRYFNSKVPYG 262
QY 186 -----QTSDDGKKYRTRAQGHASTAVGAMSYAAGHFSNAGFTGYATAEAAVSLAVGL 238
DB 263 FNYADNNTITDDKYD-----EQHGM-----HVAGIIGANGTGDPAKSVGV 305
QY 239 AAAA-----TKQSSIAGSNAKANAFATAIGGNVTVNLRGVALFGSGLDRDN 289
DB 306 APEAQLLAMKVFNSDTSKTSATVWSAIEDSAKIGADVLNMSLG--SNSGNQTL----- 360
QY 290 NTDSAYVPLGTLADQYKATPQGSTDIFSGNSNNNSIRRKIINVAGSRDTPAVN 349
DB 361 -----DPELAAVQANESGTAA-VISAGNSCTSGSA-----TEGVN 395
QY 350 VAQLKLVEELANKRTIFKGDGN-----NSNSVERGLGNTLTITKGAQTWALTEANIGVVD 406
DB 396 K-----DIYGLQDNEMVSGPSGTSG-----ATTVASAENTDVIQA--VTITD 436
QY 407 GNLKVKLAKELTGLTSVSTANKITVSNWNNNAELQSGLTFSPITGKTDKTVYSIDG 466
DB 437 GTGL--QLGPETIQLSS-----HDFGSGFDOKREYIV-- 466

QY 467 LKFTNDSNSIATKGT---TRITKKKIGFAGTND-GVDESKPYLDNEKLVKGVNSTLNGS 521
DB 467 ---KDSGNLSKALADYTADAKGKIALVRGEFSDKKQYA-----QAAGAAG 513
QY 522 LRVNNTTGNKQIQVANGIKPATVANNVNTSA-TVCTARITEKI-----GFAGTNDG 574
DB 514 LLIIVNTDGT-----ATPMTSIALTTTFFGLSSVYTGOKLVWDVWTAIPDDSLG 561
QY 575 VDEQAPYLDKERL-----KVGREITTDGSIAGNHNHKITGLTINGIANTDAVTI 622
DB 562 VKITLAMPNQKYTEDKMSDFTSYGVSNI.SPKDITAPGGNINWSTNNNGYTNMSTGSM 621
QY 623 KOLKDAKPTLNAGDGI---SINSNGDLVDSSNGNTTITTYNISVKTTLNS-----NGTSG 675
DB 622 -----ASPFIAQSQAALLKQALNNKNPNFYAYYKQLKGTALTDFLKTVMENMTAQPINDINY 676
QY 676 NKFVSNAHDNNSLVTAKDLADYLKNVNET--ADSALPSPKPVQNGDNSNNAITVCKDTN 733
DB 677 NNVI-VSPRRQAGLVDRKAAIDALEKNPSTVVAENGYPVELKDTFTSD----- 725
QY 734 GKTFTNLKLGKGVNITTNRATGTVTFGIDOSNGLTTPKLTIVGSDTNCNRLVIEQVPSA 793
DB 726 -KTF-----TFTNRTHLTYQMD-SNTDTNAVYTSATDPNSGVLYDKKIDGA 773
QY 794 DGNSTNINIKLSPTLPSIASPSGRNIALGNTIE-EKDKSNAASIDVDLN-----AGEN 846
DB 774 -----AIKAGS-----NITVPAGKTAQIEFTLSLPKSFDOQFVGEFLNEKSGDSRLN 822
QY 847 LK-----NNGKDKDFVSTYDVFIDGNATTAIVTDEANQTSKVAYDVNVDEKTTIE 898
DB 823 LPYMGFFGDNDCK-----IVDSLNG-----ITYSPAG----- 850
QY 899 LTGONG-----KKQGVKIKLTETSTNGNATFTSDDDHALVKASDIAGNLNTLAEI 952
DB 851 --GNFGTVPLLNKNTQTYGGMVTDADGNKV---DQAIATFSSDKNALYDISMKY 904
QY 953 HTTGTGTALOTFTVRKVDENDKADDTNATITVGKDTGSKVNTLK----- 998
DB 905 YLLRNISNVQDIL-----DQGNKVTTLSSSTNRKKTYYNAHS 943
QY 999 -----LKGKNGLDIKTDKGTVTFTGINTOSGLKAG----- 1028
DB 944 QOYIYVNAWPGTYDQORDG-NIKTADGSIYTRI---SGVPEGGKRQVDFVPFKLDS 999
QY 1029 DSTLLNNGLSIKN-----TASNEIQVAGDGVKAFMNVNGV-----CAG 1069
DB 1000 RAPTVRHVALSAKTENGKTQYLLTAEAKDDLGLDQATKSVKTEINEVTNLDATFDAGTT 1059
QY 1070 IDGTTRI-----TRDEIGFTGTNGSLDKSKPHLSKDG-----INAGG 1106
DB 1060 ADGYTKIETPLSDEQAQALGNGSDNSAELYLDNANATDQDASVOKPGSTSFDLIVNGCG 1119
QY 1107 --KKTNIQSGEIAKNSHDAVT-GGKIYDLKTELENKISSATAQNSLHFEFSA-DEQG 1162
DB 1120 IPDKISSITTG-----YEANTQGGTYTFSCTYPAADVGTVYDQAQKKHDLNTTYDAAT 1173
QY 1163 NNFTVSNPSSYSDTSKTSVITTFAGENGIT---TTKVNKGVVRVICIDOTKGTTPKLT-V 1217
DB 1174 NSFTASMPVTNADYAAQVDLYADKAHTQLKLFHFDTKVR-----LMAFTFDL 1220
QY 1218 GNNNGK-----GIV---INSONQNTITGL-----SNTLIANTYN 1248
DB 1221 KPNNGSDQTSSEATIKVTGVSADTKTVNVGHVVAALDAQHFEFSDVPVNYGONTIKVTAT 1280
QY 1249 DKGSVRTTEOGNIKDEKTRAASIVDVL.SAGENLOGNCEAVDFVSTYDT-VNFANG--N 1305
DB 1281 DKDGNTEQKTISSYDP-----DMLKKS-----TFDQGVKFGTNRFN 1320
QY 1306 TTTAKVYDDTSKTSKVVDVNVDDTTIEVKKKLGVK-----TTTLTSTGTGKANKFAL- 1359
DB 1321 ATSAKF-YDPKGTGIATITGVKKHPTTLQVDGKQIPIKDDLTFSFTLDLGTGQKPFVGV 1379
QY 1360 -----SNQATGDALVKASDIVAHLNLTSGDIQTAKGASQANNSAGYVDAGNKKVIYDSTD 1414


```

Db 1380 VGDTONKTFQALSFILDAVA--PTLSLDSST-----DAPVYNDPNFOITGTATYD 1429
QY 1415 NKYYQAKNDGTVDKTEKAKDLVAQAQTPDGTLAQMNVKSWINK-EQVNDANKKQG---- 1470
Db 1430 NAQYLS-----LSINGSSVASQVEDININSKPGHMA 1461
QY 1471 INEDNAFYKG---LEKAASDNKTKNAAVTGVDLNAVAQTPLTFAGDGTGTAKKLGETLTI 1527
Db 1462 IDQPKLLKGGKVLAVAVTDSNDN---TTTKNITVYEPKKTLAAPTTPS-----TT 1511
QY 1528 KGGQTDNKLTDNNIGVAGTGTGFKLAKDLTNLSNVNAGG---TKIDEKGISFVDANG 1584
Db 1512 EPAQIVT--LTAN-----AAATGETVQVSAD-----GKTYQDVPAAGVT--ITANG 1554
QY 1585 QAKATPVLASANGLDLGGKVIISVNGKGTDTDAANVQQL-NEVRNLILGLGNADGNNOVN 1643
Db 1555 TFKFKSTDLYGNEPAVDYVVTNI-----KADDPQAQLQAQKELTNL----- 1596
QY 1644 IADIKDPNPGSSSNETVTKAGTVLGGKGNNDTEKATG--GVQGVGVDKDXNANGDLSNV 1701
Db 1597 IASAKTSLASGKYDDATT-----TALAAA-----TQKAQTALDQTNASVDSLGTGANRDLQT- 1647
QY 1702 WKTKQKDGSKKALLATYNAGQNTVYNNPAEAIDRINEQIRFHVNDGNOEPVVOGRN 1761
Db 1648 -----AINQLAAKLPAKKTSL-----NQLQSVKDAL 1675
QY 1762 GIDSSASGRKHSVAIGFOAKADGEEAAVAIGROTQAGNQSIAIGNAQATGDQSTAGTGNV 1821
Db 1676 GTDLGNQTDPSGKTFTAALDLVA-----QAQAGTQT--DDQLQAT--LAKILDEV 1723
QY 1822 VAKKISGALGDFSTVKADNSYSGVNNQTFDTATDQTFVGVGNITVTSNSVALGNS--- 1878
Db 1724 LAKLABG-----IKAATPAEYGNAK-----DAATGKTWYADIADTLTSCQASADSKLA 1773
QY 1879 -----SAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAGQTAGVAVSVG---A 1926
Db 1774 HLOALQSLTKVAANAVEAAKTVKGGDGTGTGSDKGGGQGTAPAPAGDTGKDKGDEGSQPS 1833
QY 1927 SGAERRIQNAAGEVSATST 1946
Db 1834 SGG-----NIPKPTATTST 1848

RESULT 12
CBPA.CLOCL
ID CBPA.CLOCL STANDARD; PRT; 1848 AA.
AC P38058;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULOSE BINDING PROTEIN A PRECURSOR.
GN CBPA.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92228810; PubMed=1565642;
RA Shoykov O., Takagi M., Goldstein M.A., Doi R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
binding protein A.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC -!- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE
ENZYMES.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73817; AAA23218.1; -.
DR PIR: A44140; A44140.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; -.
DR InterPro: IPR002102; -.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00963; Cohesin; 9.
KW Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 1848
FT DOMAIN 29 189
FT SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;
SQ
Query Match 4.0%; Score 410.5; DB 1; Length 1848;
Best Local Similarity 21.3%; Pred. No. 8.e-08;
Matches 443; Conservative 252; Mismatches 786; Indels 603; Gaps 103;
QY 181 ILKKIQTS-----TDGKIKYRRT--AOG-----HASTAVGAMSVAQGHFSNAFGTYA 226
Db 52 IIKITNTSDSLNLDNVKRYYYTSDGTQGTQCFWCDHAGALLG-----NSYDNT 101
QY 227 TAAEAYSLAVGAAQAATKQSSIAVGSNAKANAPA---ATAIGGTVVNLGRGVALGFGSQ 283
Db 102 SKVTANFVKETASPTSYDTYVEFG-----FASGRATILKKGFITIOGRITKSDWSNY 154
QY 284 ILDRNNTDASAVPLGKTLADQY---KATROGSDTDIFSIGNSNNSNNSIRKKNV 339
Db 155 TQTNDYSFDSASSTPVYVNPKNVGTGIGGAKVLGTAPGDPVS-----SIINPT 201
QY 340 AGSRDQDAVNAOALKVEELANRKITPKGDGNNNSVERGL-----GNLTITK----- 388
Db 202 SATFDKNTVKQADVKTTMTLNGN--TFKITTDANGTALNASTDYSGNDVTISKAYLAK 259
QY 389 -----GDAQTNALT-----BANIGVYTDGNGLKVYKLAKELT-----G 420
Db 260 QSVGTTTLNPNFSAGNPQKLVITVVDTPVEAVTATIGKVQVNAGETVAVPVNLTKVPAAG 319
QY 421 LTVSVATNKITVSTNTNNRAELQSGGLTSP---ITGKTDKTVYSI----- 464
Db 320 LATIEL--PLTFDSASLEVYSITAGDIVLNPVSFSTVSGSTIKLLFLDDTLGSLQITK 377
QY 465 DGLKETNDSNSIATKGT--RITKKIGFAGTNDGVDKESPYLDNEKLVKGNSTLNSGLT 523
Db 378 DGVFAITTFKAKAITGTTAKVTSVKL--AGT-----PVGDAQLQEKPCAVNPGTGT 427
QY 524 VNNTTGNKQIQVGANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDQEQAPYLD 583
Db 428 INFIDNRMQISVGTATVK-----AGEIAAVPVTLTSPST---GIATAEAQVSFDATLLE 479
QY 584 KERLKYGRVEITTDSCIGNAGNHKITLGTNGIANTDAVTTIKLKDAKPT--INAGDGSINS 642
Db 480 VASVTAG--DIVLNPTVN-----FSYTVANGNVIKLLFLDDTLGSLISKDGVFTI 528
QY 643 N-NGLDVDSGNGNITPTTYNISVKTTKLNSNCTSG--NNKFSVSNNAHDNNSLVTAKLADY 699
Db 529 NFKAKAVTST--VTTP---VTVSGTPVFADGTLAEVQSKTAAGSVTIN---IGDPILEPT 580
QY 700 LNKVNETADSAFSPKQVQNGDNNNAITVGKDTNGKTFNTFLKLGENGVTNITNRTATG-- 757
Db 581 ISPVTATFEDKKAPA-----DVATMTLNGVTFN-----GITGLTSDYSISGNV 624
QY 758 -----TVTFGIDQSNGLTTPKL-----TVGSDT--NGNRLVI 787
Db 625 VKTSQAYLAKQPVGDLTLTFNFSNGNKRTATAKLAVSVSIKDAKPTVTATVGTATVAGETVA 684
QY 788 EQV--PSADGNSPKNIKIGLSPTLPSPASPSGRNIALGNTIEEKDKSNAASI-----DDV 840
Db 685 VPTLSNVSGISIAELQSLFSDATLLLEVSVITAGDIVLNPVSFSSVNVNGSTIKLLFLDDT 744

```


1760 RINGDSSASGKHSVAIG---FOAKADGEAAVAIGROTQAGNSIAIGDNRQAOTGDOSIAI 1816
1653 -KCYDTYISGS-TVTISKAVLATLADGSATL-----BEVFNQGSASAKLRUTIV- 1698
1817 GTGNVAGKHSALGDP-----STVKADNSVSGVNNQFTDATOTDVFVGNNTITVTESNS 1872
1699 -----PAVVDPPVTDFAVKIDKVSAAAGSTVKVPVSLINYSKVG-NVCVAE-YK 1745
1873 VALGNSAISAGTHAGTQAKKS-----DGTAGT---TTTACATGTVKVG 1912
1746 ISFSSVLTYVTAGTTSIKNPVNFSSQLNGNTITLLFFDNTIGNELLITADGQFATI-- 1803
1913 FAGQTAVGAVSVGASGAERRIQNVA-----AGEVATSTDAVNGS 1952
1804 ---EFKNAATSGTTAEVKVATISSFADASLTETITKVATVNGS 1844

RESULT 13

HLIA_SERMA STANDARD: PRT: 1608 AA.
ID HLIA_SERMA
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HEMOLYSIN PRECURSOR.
GN SHLA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RX STRAIN=SN8;
RC MEDLINE=88257037; PubMed=3290200;
RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia
J. Bacteriol. 170:3177-3188(1988).
RL - FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
DEFINED.
CC - FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
REQUIRES SHLB FUNCTION.
CC - SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC - SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M22618; AAA50323.1; --
DR PIR; A28182; A28182.
KW Hemolysis; Toxin; Outer membrane; Signal.
FT SIGNAL 1 30
FT CHAIN 31 1608 HEMOLYSIN.
FT SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;
SQ

Query Match 3.9%; Score 404; DB 1; Length 1608;
Best Local Similarity 20.1%; Pred. No. 1.3e-07;
Matches 398; Conservative 255; Mismatches 702; Indels 624; Gaps 97;

QY 219 SNAFGTYATAEAYSILAVGLAAQAQTSIAVGSNAKANAFAAIGTNTVNL-----GR 274
Db 3 NNNFRLSAGKLAALAILAASAGAYAEIIVAANGANGPGVSTAATGAQVVDIVAPNGN 62
QY 275 GVA-----LFGSOLDRDNDNTDASAVVPLGKTL- 303

841 LNAFNLNK-----NGKDKDFSTYDITVDIDGNATTATVTYDEANOTSKVAYDVNDE 894
Db 745 LGSOLISKDGVFATINFAKSVSTVTPVKVSGTVPFADGTLAE-----LSYETVAGS 798
QY 895 KTIELTDGNGKQLGVKTIKLTETSTNGNATTTSTDDDHIALVKASDIAGNLN----- 946
Db 799 VTINAIGP-----VKTV-----TATVGTATVKSGETVAVPVTLSNVPGIATAELQLSFD 847
QY 947 -TLAEIHTTKG-----TANTALQTEVVKVVDENDKADDT-NAITVKGKDGTSKVNVT 996
Db 848 ATLEVASITVGDIVLNPNSVNFSSVNGSTIKLL-----FLDDTLGSLISKDGV---LAT 900
QY 997 LKLGKNGLDIKTDK---DGTVTETGINTQSLKA-----GDST 1031
Db 901 INFKAQTVTSTVTPVAVSGTVPFADGTLAEQLSKTVAGSVTIEPSPQVKVTATVGTAT 960
QY 1032 TLNNGLSIKWTASN-----EQIQVGADG--VKFAMVNGVVGAGIDGTRITRDEIGF 1083
Db 961 VKSGETVAVPVTLSNVPGIATAELQVGFADTLLEVASITVGD-----VLNPSVNF 1011
QY 1084 TG-TNGSLDK-----SKPHLSKDG-----INAGGKKITNIQSGEIAKNSHDAVTGCKI 1130
Db 1012 SSVNGSTIKLLFLDDTLGSLISKDGVLATINEFAKTVTSKVTTPVAVSGTVPFADGTL 1071
QY 1131 YDLKTELENKISSTAKTAQNSLHIESVADEOG-----NNFTVSNPYSSYDTSKTSDVITF 1185
Db 1072 ----AELNKKTVAGSVTIEPSPQVKVTATVGTATVKSGETVAVP-----VTL 1115
QY 1186 AGENGITTKVNGVVRVGDQTKGLTTPKLTIVGNNGKGIIVN-SONGQNTITGLSNTLA 1244
Db 1116 SNVPGIAT-----AELOVGFADTL-LEVASITVGD-----IVLNPNSVNFSSVNGSTIKLL 1165
QY 1245 NVTNDKSVRTVEQGNIIKDEKTRAASIVDVLNAGFNLOGNGEAVDFVSYDVTNFPANG 1304
Db 1166 FLDDTLGSLISKDG-----VLA-----TINF-KA 1189
QY 1305 NTTAKVT-----YDTSKTSKVYDVNVDDTTIEYKDKLGVKKTTLTSTG-- 1351
Db 1190 KTVTSKVTTPVAVSGTVPFADGT--LAELKYETVAGSVTIEPSPQVKVTATVGTATGKV 1247
QY 1352 --TGANKFALSN-----QATGDA-LVKASDIVAHLNLTLSGDIOTAKGASOANNAG 1399
Db 1248 GETVAVPVTLSNVPGIATAELQVGFADTLLEVASITA-----GDIV---LNPSVNFSS 1297
QY 1400 YVDADGNKVIY-DSTONKYIOAKNDG---TVD-KTRKVAADKL--VAQAQTP---DGTILA 1449
Db 1298 VVNGSTIKILLDDTLGSLISK-DGVFAINFKIKAVPSTGTPVAVISGTPVFAADGTLA 1356
QY 1450 QMNYKSVIN-----KEOVNDANKQG-----INEDNAFVKLEKAASDN 1488
Db 1357 EVQYKTVAGSVTTAAADIKAVKATVGTATGKAGDTVAVPVTLSNVSGIATVELQLSFDAT 1416
QY 1489 KTRKNAVTGDLNVAQTPLETFAG-DTGTTAK-----KLGETITIKGGQTD- -NKL 1537
Db 1417 LLEVASITAGDI--VLNPSVNFSSVNGSTIKILLDDTLGSLISKDGVFATVNFVKAS 1474
QY 1538 TDNNTIGV---VAGT---DGFTVKLAKD-----LTLNLSVNAVGGTKIDKGLSFDVANG 1584
Db 1475 TATNSAVTPTVSGTVPFADGTLAEKLSESAAGRLTILPTVIVDSTVAPTATVTFKAKN 1534
QY 1585 QAKANTPVLSANGLDGLGKVISNVKGTGKTDAAANVOQLNEVRNLLGLGNDNADGNQVNI 1644
Db 1535 ADAALITMTLNGTTFSAIKNGTATLVKGTDTYVSENVTISK----- 1575
QY 1645 ADIKKDPNPGSSSNRTVIAKGTV-----LGGKGNNDTEKLTATGGVQGVVDKGNANGDLSN 1700
Db 1576 AVLAK-----QTGVTTLFEFVDFKNGSAKVAVVAVKEIQI-----VNSTITP 1615
QY 1701 VVVTKQDGSKKA-LLAYNAGQNTVYNNPBAIDRINEQGIREFHVDNGOEPVVOG 1759
Db 1616 VVATFEKTAAKQADVVTVMSLNGNT-----FSAIKNGTTLV--- 1652

Db 63 GLSHNOYQDNVNPQGNVLNNSREAGLSQLAGQIGANPNLGGREASVILNEVGRNPSLL 122
QY 304 -----ADQYKATROGSDTDIFSIGNNNNSIRKIIINVGA-----GSRDTD 346
Db 123 HGQOEIFGMAADYVLANPNCISQSCGFINTSHSLVWGNPLVENGVLQCYSTFGNRTL 182
QY 347 AVN-VAQLKVLVELANRKITFKGD-----GDNNSNSVERGLG-WFLTKIGDAQTN 394
Db 183 SLNGLTNAGGVLDLAPKIDSRGEVIVQDPKQNGKVTSAAINAISGLNKNRVARDGTVOAS 242
QY 395 -----ALTEANIGVYTDGNGKLKAKELTG-----LTSVSATNKIT 431
Db 243 QOMPTALDSYLSGMOAGRINIINTAGSGVKLAGSLNAGDELKVKAYDIRSERVDDAS 302
QY 432 VSNNTNNNAELQSGGUTFPITGKTDKTVYSIDGLKFTNDSNSIATGTRITKKKIGF 491
Db 303 SNKNGGDNYQNYRGGI-----YVNDRSSQTLTRTELKGNISL 341
QY 492 AGTNDGVDESKPYLDNEKLVGNSTLNSGSLTVNN-----TTG 529
Db 342 VADNHA-----HLTATDINGEDITLOGGKLTLDGQOLKOTQGTDDRWFYSWQYDVIRE 395
QY 530 NKQIQVANGANGIKATVANNVANTSATGTARI--TEEKIGFACTNDGVDEQ----- 578
Db 396 REQLQ-----QAGSTVAASGSAKLISTQEDVKLLGANVSADRALSVAARDV 442
QY 579 --APYLDKERLKYGRVEITTDGINSAGNHKITGLTINGIANTD---AVTIKQLK-DAKPTL 632
Db 443 HLAGLVEKDK-----SSERCYQR-NHTSSLRGTGWSNDESELSKASELSEGETLT 493
QY 633 NAGDGSINSNGDLVDSSGNIITPYNISVKTILNSGTSNNKFSVSNADHNSLVT 692
Db 494 KAGRNV---STQAKVHAQRDLT-----IDADNQIQGVQVKTANAKAV 533
QY 693 AKOLADYLNKVNETADSALPSFKVQMGDNSSNA---ITVGKD-TNGKPTNTLKLKENG 747
Db 534 RDKTSWGG-----IGGDNKNNSNRREISHASELTSG---GTLRLNGQQG 576
QY 748 VNITTRATGTVTFGIDQSNGLTTPKLTIVGSDTNGRLVIEQVPSADGNSTKNIKGLSP 807
Db 577 VTITGSKARG-----QKGEVTA-----THGG-----LRIDNALST 607
QY 808 TLPASIPSPGNTALGNTIEEKDSNNAASIDVNLNAGFNLK-NNGKDKDFV-STYDITVDF 865
Db 608 TVDKIDARTATPNTITSSSHKADNSQSTASELSKSDTNLTLSHKDADVIGSQVASGGE 667
QY 866 IDGNATTATVYDEANQTSKAVYDVNVDEKTIETLDGNGKKQLGVKTIK--LLETSTNGN 923
Db 668 LSVESKTNINVKAAERQ-----NIDEQKALTVNGYAKEAGDKQYRAGLRIEHTRDS 721
QY 924 ATTFSTDDHALVKASDIAGNLNLAEEIHTTKGTANTALQTFVTKVDENDKADDTNAI 983
Db 722 EKTTRTEN-----SASSLSG----- 736
QY 984 TVGKDGTSKVNLTLLKLGK-----NGLDIKDKGTVTFGINTQSGLKAGDSTTLNNGL 1038
Db 737 -----GSKVLKAEKDVTFSGSKLVADK-GDASVSGNKVSVSLAADDKTSAN--- 780
QY 1039 SIKNTASNEQIQVADGVKPFAMVNGVVGAGIDGTTTRITRDEIGFTGTNGSLDKSKPHLS 1098
Db 781 -----TEQTKIGG-----GFFYT-GGIDK-----L 799
QY 1099 KDGINAGKKITNIQSGEIAKNSHDAVTGKIIYDLKTELENKISSIAK-TAQNLSHEFSV 1157
Db 800 GSGVEAG-----YENKTOAQASKAITSGS--DVAGNL--TINARDKLTQOQAQH--SV 847
QY 1158 ADQGNFTVSNYSYDTSKTSKSDVITFAGENGITITKVNKGV-VRVGIDOTKGLTTPKLT 1216
Db 848 GGAYOENAGVVDHLAAADTASTT-----TTKTDGVGNIGANVDYS-AVTRP--- 892
QY 1217 VGNNGKGIWINSONGONTITGLSNTLANVTNDKGSVRTTEQCNITTKDEKTRAASITDV 1276
Db 893 VERAVGAAKLDTATGVINDIGGIG--APNVGLDIGA-----QGSSEKSSSSQAVVSSV 945

QY 1277 LSAGFNLOQNGEAVDFVSTYD---TVNFA--NGNTTTAKVTYDDTSKTSK-----VYYD 1325
Db 946 QAGSIDINAKGEVRDOGTQOASKGAVNLTAADSHRSEAAANRODEQSRDTRGSAGVRVT 1005
QY 1326 VNYDDTTIEVKDKLGVKKTITLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDI 1385
Db 1006 TTGSDTLTVDAK-----GEGGTQ--RSNSSASQA-----VTGSI 1036
QY 1386 QTAGASQANNSAGYVDADGNK-VIYDST---DNKYFQAOKNDGTVDTKTKVAKDKLVAQA 1441
Db 1037 DAANG-----INVVKDAIYOGTALNGRGKTAVNAGG-DIRLOQASDK---QS 1082
QY 1442 QTPDGLTAQNNVSVINKEQVNDANKKQGINEDNA-FVKGLEKAASDNKTKNAAVTVGDL 1500
Db 1083 ESRSG---FNVK-----ASAKGFTADSKNFGAGFGGTHNGESSSTAQVGN 1128
QY 1501 NAVAQPLTLAGDTGTAKKLGETLTIGK---GOTDNKLTNNIGVAGTGDFTVKLA 1556
Db 1129 -----SGQGVVEL-RAGRDLTLOGTDVKSQGDVLSAGNKVALQAAESTQTRKES 1177
QY 1557 KDLTNLNSVNAGGKIDKEGIFSVDANGQ---AKANTPVLISANGLDLG--GKV-ISNVGK 1610
Db 1178 KLSGND-LGAGSDSKETGGNLSAGGAFDIAKNESATERQGATTASDGKVTLSANGK 1236
QY 1611 GTKDTPDAANYQ-----QLNEVRN-----LLGLGNDNADGNQVNI 1645
Db 1237 G---DDALHLQAGKAVSGGSAALEAKNGGILLESAKNEQHKDNWSLGI-KANAKGGQTFNK 1292
QY 1646 DI--KDDPNSGSSSNRTVIKAGTVLGKGNNDTEKLATG-----GVQV 1686
Db 1293 DAGGKVDPNFGKDPH--TLGAGLKVGEQDDKTHANTGITAGDVTLSNGKDTRLAGARV 1350
QY 1687 GVDK-DCGNANGDLSNVVKTOKD---GSKKALLATYNAAGTOTVYTNPAEAI-DRINEQ 1741
Db 1351 DADSVQGVKGGDLH---VESRKDVENGKVDV-----DAGLSH--SNDPGSSITSKLSKV 1400
QY 1742 GIRFFHVNDGN-DEPVVQGRNGIDSSASGKH-SVAIGFOAKADGAAVAIGR----- 1791
Db 1401 GTPRY---ACKVREKLEAGYKVVADATTDKYNSVARRLDPQDDTTGAVSF-SKABGKVTL 1457
QY 1792 QTAGNQSIA-IGDNAGATGDSIATIGTGNVAGKHGSAIGDPTVKADNSYSVGNNNQF 1850
Db 1458 ATPAGEKQGPLMDRGART-----VGGAVKDSITGPAGROGHLKV--NADVYNNN-- 1505
QY 1851 TDATQTDVFGVGNITVTESNVAL--GSNSAISAGTHAGTQAKKSDG----- 1896
Db 1506 -----AVGESATAGKNVALQVGGQTLTGGEIRSQGKVELGGSOVSQDYNQ 1556
QY 1897 --TAGTTTTAGAT--GTVKGFAGQTAGVAVSVGASGAERRIQNVAAAGEVSATSDAYNG 1951
Db 1557 RYQGGGRVDAATAVGGLLGGAQKQSVAGNVP-----FASGHASTQQADAKAG 1603

RESULT 14

FHAB_BORPE
ID FHAB_BORPE STANDARD; PRT; 3591 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE FILAMENTOUS HEMAGGLUTININ.
GN FHAB.
OS Bordetella pertussis.
OC Bacteria: Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
haemagglutinin: a protein processed from an unusually large

Db 1640 RNDIALDVADFTNTGSLYAEHDAITLTAQGTQBDLVDDHILPVAEGTLRVKAKSL--T 1697

QY 1525 LTIKGGQ-----TDNKLFDNNIGVAGTGDFTVLAQKDLT-----NL-NSVNA----- 1567

Db 1698 TEIETGPGSLIAEQVINDKQAIIVG-----KDLTSSAHGNVANEANALLWAA 1748

QY 1568 -----GGTKIDEGKISFYVDANQAKANTPVLSANGL-----DLGKVISNVGKGTK 1613

Db 1749 GELTVKAQNTITKRAALIEAGGNARLTAVALLNKLGIRAGEDMHLDPRIENTAKLSG 1808

QY 1614 DTDAAANVQOL--NEVRNLGLGNDN---ADGNQVNIADIKKDPNSGS--SSNRTVTKAGPV 1667

Db 1809 EVQRKGVQDVGGEGHCRWSSGIVYVWLRAGNGKKAAGTAAAPWYGGDLTAEGSLIEVGKD 1868

QY 1668 L-----GGKGNNDTEKATGGVGVGVDKNGANDLSN----- 1700

Db 1869 LVLNAGARKDEHRLHLLNEGVIQAG--GHGIGGDVDRNSVWRTVSAMEYFKTPLPVSLTA 1926

QY 1701 -----VW-----VKTQD----- 1708

Db 1927 LONRAGLSPATWNFOSTYELLDYLLDQNRYEYIWLGYPTTWSVNTLKNLDLGYQAKPA 1986

QY 1709 -----GSKKALLATYNAAGQTNVYNNPAEADIRINE 1740

Db 1987 PTAPPMPKAPELDLRCHTLESAGRKIFGEYKKLQGEYKAKMAVQAVAYGEATRRVHD 2046

QY 1741 Q-GIRFFHYNDQNEPVGVRNGIDSSASGKHSVAI--GFQAKADGEA----- 1785

Db 2047 QLGQRYKALGMDAETKEVDGLIQEFAADLRTVYAKQADQATIDAETDKVAQRYKSQID 2106

QY 1786 -----AVAILGROTAGNQSIAGIDNNAQTGDOSIAGTGNVYVAGKHSAGIDPSTVKADN 1840

Db 2107 AVRLQAIQPGRYTLAKALSAAALGADWRALGHLSQLMQWRKDFKAGKRGAEIA----- 2157

QY 1841 SYSVGNNOFTDATQDVFVGNNITVTESNVALGNSAISAGTHAGTQ--AKKSDGTA 1898

Db 2158 -----FYPEQET-VLAAGAGLTLS-NGAIHNGENAAQNGRPEGKIGAHSAITSVS 2206

QY 1899 GTTTTAGATGTGKGFAGQAVGAVSVG-----ASGAERRIONVAAGEV-----SATST 1946

Db 2207 GSFDAIRDVGLKRLDIDDAALAVLNPHIFTRIGAQTSLADGAAGPALAQARQAPET 2266

QY 1947 DAV-----NGSOLYKATQSIANATNLDHRIHQENKANAGIS--SAMAMASNPQAYIQC 1999

Db 2267 DGMVDARGLSADALASLASDLAAGQLEVSRRNAQVADAGLAGPSAAPAAGVADV-G 2325

QY 2000 RSMVTGGIATHNGQGAVALGSLK 2022

Db 2326 VERVTG---DQVDPQVAVGLEQ 2345

RESULT 15

P2P_LACPA

ID P2P_LACPA STANDARD; PRT; 1902 AA.

AC Q02470:

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-RT)

GN PRT.

OS Lactobacillus paracasei.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OC Lactobacillus.

OX NCBI_TaxID=1597;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCDO 151;

RX MEDLINE=92381481; PubMed=1512565;

RA Holck A., Naes H.;

RT "Cloning, sequencing and expression of the gene encoding the cell-envelope-associated proteinase from Lactobacillus paracasei subsp. paracasei NCDO 151.";

J. Gen. Microbiol. 138:1353-1364(1992).

[2]

SEQUENCE OF 189-196.

MEDLINE=92226694; PubMed=1564442;

Naes H., Nissen-Meyer J.;

"Purification and N-terminal amino acid sequence determination of the cell-wall-bound proteinase from Lactobacillus paracasei subsp. paracasei.";

RT J. Gen. Microbiol. 138:313-318(1992).

RL J. Gen. Microbiol. 138:313-318(1992).

CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE GROWTH OF THE BACTERIA ON MILK.

CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED, E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS, ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED INSULIN B-CHAIN.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE SUBTILASE FAMILY.

CC -----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC -----

EMBL; M83946; AAA25248.1; ..

DR PIR; B44858; B44858.

DR HSP; Q99405; IMPT.

DR MEROPS; S08.019; ..

DR InterPro; IPR000209; ..

DR InterPro; IPR001899; ..

DR Pfam; PF00746; Gram_pos_anchor; 1.

DR Pfam; PF00082; Peptidase_S8; 3.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE_ASP; 1.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.

KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;

KW Transmembrane.

FT SIGNAL 1 33 POTENTIAL.

FT PROPEP 34 187 POTENTIAL.

FT CHAIN 188 1902 PII-TYPE PROTEINASE.

FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).

FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).

FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE PROTEINS.

FT SEQUENCE 1902 AA; 200253 MW; D8C9F38CEE5DA582 CRC64;

Query Match 3.9%; Score 399; DB 1; Length 1902;

Best Local Similarity 20.7%; Pred. No. 2.3e-07;

Matches 450; Conservative 249; Mismatches 880; Indels 592; Gaps 100;

QY 31 GSCATQGVGSVRTLSF-----ARIATLAVLVIGAT-----LNGSAYAQOITTK-I 74

Db 19 GALAVLPVGEIOAKAISQOTKYVSSLANVTYKAAQATDTTAAATNQATLAAKGI 78

QY 75 EIGQTNKI-----NNTLKGDALATGEASIAFGSLKSAQGSIAI 114

Db 79 DYNNKLNKVOQDQTYVDVIVQMSAAPASENGTLRTDYSSTAEIQOETNKVIAAQAQSVKAAV 138

QY 115 GSVKPPDPNNGSNGV-----GSHAKGNESIAI- 141

Db 139 EQVTQQTAGESYGVVNGFSTKVRVVDIPKLQIAGVKVTLAKVYPTDAKANSMAVQ 198

QY 142 -----GGDYLAEGBASIASDLYLPKNLDLKNFHLIHGHEILKIKIOTSTDG 191
Db 199 AWMNXYKKGEGTVVSDIGDIPDHKMLRSLDDKOVK-----LKYDVEKEDTDIAKHG 252
QY 192 -----KIKYRRTAAGHASTAVGAMSYAQC-HFSNAFGTYATAEAAYSYLAAGLAAQA--- 242
Db 253 RYFTSKVPYGFYANDNDTITDDTVDEQHGHHVAGIIGANGCTGDDPTKSVGVGAPEAQLL 312
QY 243 -----TKQSIAGVSNKAKNAFAATAIGGTVVNLGRGVALGFGSGLDDBDNNNTDASAY 296
Db 313 AMKVFTNSDTSATTGCSATFLVSAIEDSAKIGADVLNMSLG--SDSGNQTL----- 360
QY 297 VPLGKTADQYKATROGDSTDIESIGNNNNNNSIRKILN-----VAGASRDT 345
Db 361 ---DPEIAAVONANESGTAA-VISAGNSGTSGSATOG--VNKYDYGLODNEMVGTPTCSR 414
QY 346 DAVNVAQLKVEELANRKITPKGDCDNNNSVERGLNLTLPKGDATNALTEANIGVVT 405
Db 415 GATTVASAENT-DVISOAVTIT-DGKDLOLQGP-----TIQLSSNDEFTSGFDOKKFYVK 467
QY 406 DGNGLKVKLAKELGTLTSVSATNITVSNTNNNNELOSGLTFSPITGKTDKTVYS-- 463
Db 468 DASG---DLSSKGAADYITADAKGATIAI-----VARGELNFA-----DKOKYQA 508
QY 464 --IDGLKFTNDSNSIATGTTTIT-----KKKIG-----FAGTND--GYDESKPY 504
Db 509 AGAAGLIIVNDGTATPLTSIRLTTFEGLSSKTGOKLVDWTAHPDDSLGVKIALTL 568
QY 505 LONEKLVGNSTLNGSLTVVNTTGNKOIQVANGIKPATVANNVANTISAT-----VG 557
Db 569 LPNQYTEDKMSDEFTSYGFSVLSFKPDITAPGGNINSTQNNNGYTNNSGTSMAFPFAG 628
QY 558 TARITEEKIGFAGTNGDVEQAP-YLDERLK-----VGRVEITDSDGINAGNHKITG 609
Db 629 SALLUKQALN-----NKNPFFADYKQLGKTALTDLFKTVEANTPAQINDINY----- 676
QY 610 LPNGIANTDAVTIKOLDAKPTLNAAGDGISNNSNGDLVDSGG--NITPTYNISVKTWK 667
Db 677 --NNVIVSPRGGAGLVDDVKAIDA-----LEKNPSTVVAENGYPAVELKDFETSTDFTFK 729
QY 668 LN-SNGTSGNNKFSVNAHNNLSATKDLAD-----YLNKYNETADSLPSEKVGONGNS 722
Db 730 LFTNRTHLTYQMSNTDITNAVTSATDTPNSGVLYDKKIDGAAIKAGSDITVPAGKTA 789
QY 723 NNAITVG--KDTNGKTF--NTLKLKENG--VNITNTRATGTVTFG--IDQSNGL----- 769
Db 790 QIEFTLSLPKSPDOQFVBEFLNEFKSGSDGRNLNLYMGFFGDWMDGKIVDSLNGITYSPA 849
QY 770 -----TTPKLTVGSDTNGNRLVIEQVPSADGNTKNIILKGLSPTLPSIASPSGRNIALGN 824
Db 850 GGNYGTVPLLT--NKNTGHQYGGVMVTDADGKOTVDD-----QALAFSSDKN----- 894
QY 825 TIEBKDSNAASTDDVLNAGFLKNNKOKDFVSTYDVTDFIDGNATTATVYDEANQTS 884
Db 895 -----ALYNDLSMOYLLRN-----ISNVQ--VDILDGOGNKVTTLSSTNQOT- 935
QY 885 KVAYDVNVEKTIELTGDNKKOLGVKTKLTETSTNGNATFTSTDDHIALVKASDIAGN 944
Db 936 KTYIDAH-----SOKYIYNAPAMDG---TYIDQD-----GN 965
QY 945 LNTLAEIHTTKTANTALQTFVKKVDE-NKAD-----DTNATVVGKDGTSKGVN 995
Db 966 IK-----TADGSYTRISGVPEGGRQVDFVPFKLDSKAPTVRHVALSAKTE 1014
QY 996 TLKLGKNGLDIKTDKG-----TVTFGINTOSGLKA--GDSTTLNNGNLSIKNTASNQ 1048
Db 1015 NGKTOYLLTAEAARDLGLSDATKSVKATNEVNTLDATEFTDAGTTADGYTKIETPLSDQ 1074
QY 1049 IQVGADGVKFAWYNNGVVNGVIGDITRIIRDEIGFTGNGSLDKSKPHLSKDG--INAGG 1106
Db 1075 AQA-----LGNG-----DNSAELYLTDNASNATNODASVQKPGSTFSLIVNGGG 1119

QY 1107 --KKITNIOSGEIAKNSHDAVT--GGKIYDLKTELENKISSTAKTAONSLEHPSVA-DEOG 1162
Db 1120 IPDKISSTITG-----YEANTQGGCTYTFSGTYPAADVDTYDAQGGKHLNLTNTTYDAAT 1173
QY 1163 NNETVSNPYSSYDTSKTSDVITFAGENGI---TTKVNKGVVVRVGIDQTKGLTTPKLT-V 1217
Db 1174 NSFTASMAVNTADYAAQVDLYADKAHTQLLKHEDTKVR-----LTAPITFDL 1220
QY 1218 GNNNGKGIIVNSONGQNTITGLSNTLANVTNDKGSVRRTEQGNIIKDEKTKRAASIVDVL 1277
Db 1221 KFNG-----SDQTSATI-----KVTGTVSSD---TKTVNMGDTVAALDAQHHFSVDVPV 1268
QY 1278 SAGFNL-----QGN--GEAVDFVSTVD-----TVNFANG-----NITTTAKVY 1313
Db 1269 NYGDNITKVTATDEGNTTEQKTIITSSYDPPVLKNVTFDQGVKFGANEFNATSARF-Y 1327
QY 1314 DDTSKTSKVYDVNDDTTIEVKDKKLGK-----TTTTSTGTGANKFAL-----SNQ 1362
Db 1328 DPKTGATITCKVKKHPITTLQVDGKQISIKNDLTFSTLDLGLTGOKPGGVVVDGTTQNK 1387
QY 1363 ATGDALVKASDIVAHLNLTLSGDIOTAKGASQANNAGYVDADGKNKVIYDSTDNKYQAKN 1422
Db 1388 TFOEALTFILDAVA--PTLSLDSST-----DAPVYTNDFNFQITGTATDINAQYLS-- 1435
QY 1423 DGIWDTKTEVAKDKLYAAQAOTPDSTLAQMNKVSINKEQVNDANKK--QC1NEDNAPVK 1480
Db 1436 -----LAINGSHIVASQYADININSGKPGHMAIDOPVKLLEKN-----VLT 1476
QY 1481 LEKAASDNKTNAATVYGDNLNVAQAOTPLTFAGDTGTAKKLGTELTIKGGOTDNTKLTDN 1540
Db 1477 VAVTDSENNITTKKITV-----YVEPKKTLAAPIVTPS-----TTEPAKTVT--LTAN 1522
QY 1541 NIGVVAOTDGTVKLAKDLNLNLSVNAAG--TKIDEKGLSFVDANGQAKANTPVLSANG 1597
Db 1523 -----AAATGETVQYSAD-----GGKTYQDVPAAQVT-VTANGTTFKFKSTDLLYNE 1567
QY 1598 LDLGKGLYSNVGKGTQDTPDAANVOQLNEVRNLLGLGNDNADGNQVNIADIKKDPNSGSS 1657
Db 1568 SPADVYVVTNI-----KADDPQAQLOTAQALTNL-----IASAKTLSASGKYD 1610
QY 1658 NRTVIKAGTVLGGKGNNDTEKLTAG--GVQGVYDKDNANGDLSNVVVKTKQDKGSKALL 1715
Db 1611 DATT-----TALAAA-----TOKAQTALDQTOASVDSLTCANROLOQT-----AINOLA 1653
QY 1716 ATYNAAGOTNVYTNPAEAIIDRINEQGIREFPHVNDGHOEPVWQGRNGIDSSASGKHSVAI 1775
Db 1654 AKLPADKKTSLT-----NQLOSVKAAALGTDGLGNQTDPSGK 1689
QY 1776 GFQAKADGEAAVAIGROTQACNQSIAGIDNAQATGDSQSIAGTGNVAGKHSGAIGDPST 1835
Db 1690 TFTAALDDLVA-----QAQAGTQT---ADQLQA-----SLAKVLDVAVLAKLAEG----- 1730
QY 1836 VKADNSYSGVNNQOTDATQTDVFGVGNITVTESNSVALGSN-----SAISA 1883
Db 1731 IKAATPAEYVGNAK---DAATGKTWTADIADTLTSGQASADSKLAHLAQLOSLKTKVAA 1787
QY 1884 GTHAGTQAKKSDGTAGTTTTAGATGTGVKGFAGTAVGAVSYGASGASERRIONVAAGEV-- 1941
Db 1788 AVEAAKTAGKDDITGTSKGGGGTTPAPAGDT-----GKDKGDEGSQPSGGNIPT 1840
QY 1942 -----SATSTD 1947
Db 1841 KPATTSTSTD 1851

Search completed: September 13, 2001, 12:53:37
Job time: 847 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:55:10 ; Search time 112.99 Seconds
(without alignments)
2403.949 Million cell updates/sec

Title: US-09-361-619-9
Perfect score: 10356
Sequence: 1 MNHIYKVFINKATGTGTEMAVA.....NGSADTGHVGAAGVAGEHF 2053

Scoring table: BLOSUM62
Gapop 10.0 . Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0.

Maximum DB seq length: 2000000000

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 s
```

```
Database :
SPTREMBL_16.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_orphanelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_unclassified.*
13: sp_vertebrate.*
14: sp_virus.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	1206	11.6	2353	2	P71401	haemophilus
2	1191.5	11.5	2059	2	Q9PD50	xylella fas
3	966.5	9.3	2712	2	Q9F3X5	pasteurella
4	742	7.2	1299	2	Q9F3X6	pasteurella
5	704.5	6.8	1190	2	Q9PC04	xylella fas
6	690	6.7	1098	2	Q9C152	haemophilus
7	677	6.5	1107	2	Q9F2D8	salmonella
8	658.5	6.4	3705	2	Q9F285	versinia pe
9	591.5	5.7	2340	2	Q9ZD91	rickettsia
10	559.5	5.4	2586	5	Q9VTK8	drosofila
11	557	5.4	2106	2	Q9XC47	rickettsia
12	553	5.3	2514	2	Q9JY30	neisseria m
13	540.5	5.2	4919	2	Q9ZHL0	haemophilus
14	528.5	5.1	2021	2	Q52657	rickettsia
15	526.5	5.1	2349	2	P94750	escherichia
16	526.5	5.1	2283	2	P76347	escherichia
17	526	5.1	3029	2	Q55582	synecocyst
18	513	5.0	4152	2	Q9ZHL3	haemophilus
19	512.5	4.9	2703	2	Q9K0T0	neisseria m

ALIGNMENTS

RESULT	1	
ID	P71401	PRELIMINARY; PRT; 2353 AA.
AC	P71401;	
DT	01-FEB-1997	(TrEMBLrel. 02, Created)
DT	01-FEB-1997	(TrEMBLrel. 02, Last sequence update)
DT	01-JUN-2000	(TrEMBLrel. 14, Last annotation update)

[illegible]

Db 179 DSTLPDAVNTGVLSSSFTPNDEKTRA---ATVKDVLNAGNRIKAKTAGNVSVDL 235
Qy 142 GGDVLAEGDASIAIGSDLL-----YLPKNLDKNEFHKLHGHILK 184
Db 236 ---VSAYNNVEFITGDKNTLDVLITAKENGKTEVKTPKTSVIKEKDKLFTGKE--- 288
Qy 185 IOTSTDGKIKYRRRAOGHASTAVGAMSQAQHPFNAQFATYATAEAYSIAVGLAAQATK 244
Db 289 ---NNDTNKVT-----SNTATNDTEGNG-----LVTAKAVID-AVNKAGRWK 328
Qy 245 QSSIAVGSNAKANAPAAIPAIGNTVNLGRGVALFGSQILDRDNNIDA-----SAVPL 299
Db 329 TTT-----ANGQDPAFVASGTNVPFESGDGTAS-----VTKDNGNGITVKYDAKVG 379
Qy 300 G-----KTLADQYKATROGSDTDFISGNSNNNSIRKKIINVAGSGSRDTPAVNAQ 352
Db 380 GLKFSDKKIVADTTALTVTGK---VAETAKEDD-----KKLVNAG----- 419
Qy 353 LKLVBEELANRKTTFKGDGNN-----SNSVERGLGNTLTIKGDAQTALTEANIGVVD 406
Db 420 -DLVTALGNLSWKAKAEADTDGALEGISKQOEKAGETVTFKAG-----KNLKVQD 470
Qy 407 GNCLKVKLAKETLGLTSVATNKIIVTSNTNNNAELQSGGLTFSPITGKTDKTVISIDG 466
Db 471 GANFTYSLODALGLTS-----ITLGGTN-----GGNDAKTIVINKDG 508
Qy 467 LKFTDNSNIATGTRITRKKIGFAGTNDGVDESKPYL---DNEKLKVGNS---LN--- 518
Db 509 LTIIPAGNG-GTTGTNTISVTKDGKAGNAKITNVASGLRAYDDANFDVNLNSATDLNRH 567
Qy 519 -----SGSLTVNNTGKQIQGANGIKATVANNVANTSATVGTARITEEKIGF-AGTN 572
Db 568 VEDAYKGLLNLNEKNANKQPLV-----TDSTAATVGLDR-----KLGMWSTK 610
Qy 573 DGVDEQAPYLDK--ERLKVGRVEITTDGGINAGNHKITGLTINGIANT-----DAVTI 622
Db 611 NGTKEESNOVKQADEVLFTCAGAAATVTSKENGKHTI---TVSVAETKADCGLEKDGDTI 667
Qy 623 KOLKADKPTLNAGDGIINSNGDLVDSSGNTTPTYNISVKTITKLSNGTSGNNKFSYS 682
Db 668 KLVYDNQNT-----DNVLTVGNGTAVTKGFEI-----VKTGATDAD-----RGKVTVK 712
Qy 683 NAHONNS---LVTAKDLADYLNKVNETADSLPSPKVVONGDNS---NNALTIVKQDINGKT 736
Db 713 DATANDADKKVATVKDVATAIN-----SAAFEVKTENLTTSIDEDNPTDNGKDDALKA 765
Qy 737 FNTLKLK-GEN-----GVNLT-----TNRAATGTVTEGIDQSNGLT---TPKLTVG 777
Db 766 GDTLTFKAGKNLKVKRDKGNTIFDLAKNLEVKPAKVSDTLTIGGNTPTGGTTATPKVNIT 825
Qy 778 SDTNGNRLVIEQPSADGNSTKNT-IGLSPTL--PSIASPSGRNIALGNTIEKDKSNA 834
Db 826 STADGLNFAKE---TADASGSKNVLKGIATLTPEPSAGAKSSH---VDLNVDAKKSNA 879
Qy 835 ASIDVVLNAGFNLKNGKDKDPVSTYIDVDIGNATATVYDEANQTSKVAYDVNWDE 894
Db 880 ASIEDVLRAGMNIQNGNNDVYATYDVTNFTDSTGTFTVTYQ--KADGKGADVKIGA 937
Qy 895 KTELTGDNKKOLGVKTKLTETSTNGNATTFSTDDH-----ALVKASDI----- 941
Db 938 KTSVINDHNGKLFEG-KDLK-----DANNGATVSEDDCKDGTGLVTAKTVIDAVNKS 991
Qy 942 -----AGNLNTLAE--EIHTKGTANTALOTFTVKKVDEN----- 974
Db 992 RVTGEGATAETGATAVNAAGNAETVTSVNFKNGNATTA-----FVSKDNGINVKYDVN 1047
Qy 975 -----DK--ADDTNATIVGDKTSGV-----NTLKLKGNGL----- 1005
Db 1048 VGGLKIGDDKKIVADTTILTV-----TGGKVSVPAGANSVNNKLVNAEGLATALNLS 1103
Qy 1006 -DIKTODKGVTVTEGINTQSGLRAGDSTTLN--NGLSIRN-----TASNEQIQVGADGVKE 1058

Db 1104 WTAKADYADGESEGETDQEVKAGDKVTFKAGKNLKVQSEKDFYSLQDTLGLTSITL 1163
Qy 1059 AMVNGVVVAGIGDGTTRITRDEIGFTGTNGS---LDKSKPH---LSKDGINAGKKITNI 1112
Db 1164 GGTANGRNDTG---TVINKDGLITLANGAAAGTDASNGTISVTKDGISAGNKEITNV 1219
Qy 1113 QSG-----EIAKNSHDAVTCGGKIYDLKTELENKISSTAKTAONSL 1152
Db 1220 KSALKTYKTQNTADETQDKBFHAAVKNANEVEFVGK-----NGATVSAKTDNNGK 1270
Qy 1153 H-----EFSVADBQGNFTVSNPYSSYDTSK-----TSDVI 1183
Db 1271 HPVTIDVAEAKVGDGLEKDTGKIKLKYDNTDGNLLTVDATKGVASVAKGEFNAVTTDAT 1330
Qy 1184 TPAGENG-----ITTKVANKGVVRVIGIDQTKGLT----- 1211
Db 1331 TAGTNNANERGVVVKGSNGCATATETDKKVKATVGDVAKAINDAATFVKVENDDSATIDD 1390
Qy 1212 -----TPKLTVGNNNG 1222
Db 1391 SPTDDGANDALKAGDTLTLKAGKNLKVXRDGKNITFALANDLSVKSATVSKLSLGTNGN 1450
Qy 1223 KGIIVNSQGN-----TITGLSNTLANVNDKGSVRTTEGNIILKDEKTR 1269
Db 1451 KVNITSDTKGLNPAKDSKTDGDDANIHLNGIASTLTDLLNSCAT-TNLGGNGIITONEKR 1509
Qy 1270 AASIVDLVSGFNLQG-----NGEAVDFVSTYDVTNFAANGNTTTAKVTVYD--DTSKT 1319
Db 1510 AASVKDYLNAGNVRGVKPAASANNQVENIDFVATYDVTDFVSGDKDTSVTVESKDNKR 1569
Qy 1320 SKVYDVNVDDTTIEVKDKKLVKVTTLTSTG---TGAN-KFALSNOAT-----G 1365
Db 1570 TEV-----KIGAKTSVIKDHNGKLETKELKDANNNGVTVTETDKGEG 1613
Qy 1366 DALVKASDIVAHLNLTLSGDTQTAKGASOANNSAGYVD-----ADGNKVIVDSTDNKY 1418
Db 1614 NGLVTAKAVIDAVNKAAGRVKTT-GANGQNDDFATVASCNTVTFAGNGTTAEVT----- 1667
Qy 1419 QAKNDG--TVDKTKEVAK-----DKLVAQA---QTPDGTLAQMNKVSINKEQVINDANK 1467
Db 1668 -KANDGSIYKYVYKVAVDGLKLDGDKIVADTTVLTVADGKVTAPN-----NGDGK 1716
Qy 1468 KOGINEDNAFVKGLEKASDNKTKNAAVTVGDLNVAQTPLTFAGDTGTTAKLGLTTLI 1527
Db 1717 K-----FVDSAGLADALNKLWSA--TAGKEGTGEVDPANSAGO---EVRAGDKVTF 1763
Qy 1528 KGGQDNTNKLTDNNIGVVVAGTGTVKLAKDLTLNLSV---NAG---GTKIDEGKIS 1578
Db 1764 KAG-----DNLKIQSKDFTYSLKKELKOLTSVEEFKANGGTGSESTKITKDLGT 1814
Qy 1579 FVDANGQA-----KANTPVLANSGLDGGKVISNVGKTK-----DTDAAN 1619
Db 1815 ITPANGAGAAGANTANTISVTKDGISAGNAKAVTNVSGLKKFGGHTLANGTVADFEKH 1874
Qy 1620 VOQLNEVRNLLGLGNDN---ADGNQVNIAD-----IKKDPNSGS-----SSN 1658
Db 1875 DNAYKDLTLNDEKADNPTVADNTAAATVGLRGLGVISADKTTGEPNQEYNAQVRNAN 1934
Qy 1659 RTVIRKAGT-----VLGG-----KGNNDTEKATGG---VOVG----- 1687
Db 1935 EVKFKSGINGINVSGLTNGTRVITFELAKEGVKVSNEFTVKNADGSETNLVKGDMYYSK 1994
Qy 1688 VKDGNANGDLSNVVVKTKOK-----DGSKKALLATYNAAAGOTNYVTNNPAEAD 1736
Db 1995 EDIDPATSKPMTG---KTEKYKVENGVVYVANGSKTEVTLTNKSGS---YVTGN--QVAD 2046
Qy 1737 RINEOGRIFRFHVNDGNOEPVVOGRNIDSSAGSKHSVAIGFOAK-----ADGEA 1785
Db 2047 AIAKSGFEL-----GLADAAEAEKAFESAADKOLSKOKKAEVTVNHNKV 2090
Qy 1786 AVAIGROTQAGNOSIAGDNAQATGDO-----SIAIGTGNVYVAGHSGAIGPSTVK 1837
Db 2091 RFANGLNTRV---SAAVTVESTDANGDKVTTTFVKTDVLPLOLY---NTDANGNKIVKK 2144

Qy	1009	TDKGTVTFTGINTQSGUKAGDSTTLNNGLSIKNTASNEQIQVGADGVKFMVNGVCGA	1068
Db	999	NTDGNLTIS-----KSGDSNDV-----VFNLSKDFKVDGMTSGT--TVVNDGVKV	1042
Qy	1069	CID--GTTTRTRDEIGFTGNGSLDKSKPHLSKDGKINAGKKITNIQSGEIAKNSHDV	1125
Db	1043	GSDVALGTT-----GLTIANG-----PAVTASGIDAGSKVISHVAGAVSETSDAV	1089
Qy	1126	TGGKIYDLKTELENKISSTAKTQNSLHFSVADQGNFTVSNPYSSYDTSKTSVDITF	1185
Db	1090	NGSQLNAVQASQPVFTG-----NEGAVKRSIGQSVDIS-----GESSTAGTY	1134
Qy	1186	ACENGITTVKVKNGVVRVIGIDQTLGTTPKL--TVGNNGKGIIVINSQNGQNTITGLSNTL	1243
Db	1135	SGGN-LKSVVDPAAGRIHLQLA---DSPKFGNVVINGGK-----ISGVT---	1175
Qy	1244	ANVTNDKGSVRTTEOGNIIKDEKTRAASTIVDVLISAGFNLOGCEAVDFVSTYDTVNFA	1303
Db	1176	-----ACTEETDAV---NFSQLKSIATAVDQGWTLTASGANGSKVASGCTVDLKN	1222
Qy	1304	--GNTTAKVTYDDTSKTSKVVDVNDV-----DFTIEVKDKLGVKTTTLTSTGTG	1353
Db	1223	TGDNLTISK-----SGDSNDVFNLSKDFKVDGMTSGTTVVNDV---GVKVGSDVALGTT	1274
Qy	1354	ANKFALSNOATGDALVKASDIIVAHNTLSGDIOTAKGASQANNSAGIYDADGNKVIYDST	1413
Db	1275	GLTIANGPAVTASGIDAGSKVISHV-----AAGAVSETSDAVNGSQLNAVQOAS	1325
Qy	1414	DNKYQAKNDGTIVDKTKEVAKDKLVAQAOTPDGTLAQMNVKSVINKE-----QVNDANK	1467
Db	1326	QPVTFGTG-NEGAVKRS--LQGSVVISGESSTAGTSSGNLKSVDVEAAGTTHLQADSPK	1382
Qy	1468	KQGINEDNAPVKLEKAAASDNKTKNAATVVGDLNAVAQTPFLTAGDTGTTAKKLGELITI	1527
Db	1383	FGNVVING--CKISGVTAGTEETDAVNFSQLKSI-----STAVDOGWTLTA	1427
Qy	1528	KG-----GQDTNKLNTDNNIGVACVGDGTV--KLAKDLTNLSNVNAGGTKIDEXGI	1577
Db	1428	SGANGSKVASGGTVDLKNTDGNLTISKSGDSNDVFNLSKDFKVDGMTSGTTVVNDV	1486
Qy	1578	S-----FVD-----ANGQAKANTPV-----	1592
Db	1487	KVSSNVLLDSNELVITSHSSTSSVKTLANGESVNVNVTNVNGDGVNDDVVVNDLGLSTV	1546
Qy	1593	-----LSANGLDLGKVIKSNVKGKTDKTDAAVQOLNEVRNLLGLG---NDNADGNQV--	1642
Db	1547	GGASLTLSGINAGSHKITNVTAGTEDIDAVNFSQLKSVEADVKGWTLTASGANGSKVVS	1606
Qy	1643	NIADIKK-DPN-----SGSSNRVTI-----KAGTVLGGKGNNDTEKLATGGVQGVQDK	1690
Db	1607	GGTVDLKNTDGNLTAISKSGDSNDVFNLSKDFKVDGYTAG-----NTVNTDGVKVGSDV	1661
Qy	1691	DGNANG-----DLSNVVKTOKDGSKKALLATYNAAGOTNVVNNPA---BAIDRINEQGI	1743
Db	1662	SLUGMGLFIANGSPVITAGFPNAGOR---VISHVAVGADTDVAVNSOLKQAVQSVTVKAT	1718
Qy	1744	REFVNDGNOEPVQVRNGIDSSASGKHSVAIGFOAKADGAAVAIGRQTOA-GNQSIAI	1802
Db	1719	RYISTNDGG---TOGGNVGDGATGSKAIAAGVGTQASGEGAAVSGAAASGKGSTAI	1774
Qy	1803	GDAOATGDOOSTAIGTG-----NVVAGKHSG-----AIGDPSTVKADNSYSVG	1845
Db	1775	GRNATASADGSVALGDGAKDGGRAESYTGKISGVQNTVTGTVSVGDAKAGETRSINVA	1834
Qy	1846	NNNOFTDAT-----QT-----DVEGV--GNNTVTVES---NSVAL	1875
Db	1835	DAKAMDVNLRLQDAVAKSNLQTDMDRHEINNIEDVEKITKGDSSASSVRGMGVNMAI	1894
Qy	1876	GSNATSAGTHAGTQAKSDGTAGTTTATAGATGTVKGFAGGTAVGAVSVGASGAEERRIQ	1935
Db	1895	GTNAAVS-GTESVALGKNTNVSNADNAVATG-NGSVADRA-----NSVSVGSGGSEBQVTN	1947
Qy	1936	VAAGEVSATSTDAVNGSQLYKATQSIANATNELDHRHONENKANAGISSAMAMASMPQA	1995
Db	1948	VAAG---TADTDVAVNSQL---NOGLITAKQYTDGMVGNLRRRETSGVAAAIAATANLPOA	2001
Qy	1996	YIPGRSMVTGIAITHNGOGAVAVGLSKLSNGQWVFVKINGSADTOGHVGAAGVAGGFHF	2053
Db	2002	YVQGRGMTSVGVSYSQOGSALAVGVSAVSESGHVMVFKPFGSANTRSHVGVGAGVGYQM	2059
RESULT	3		
Q9F3X5		PRELIMINARY;	PRT: 2712 AA.
ID	Q9F3X5		
AC	Q9F3X5		
DT	01-MAR-2001	(TREMBLrel. 16, Created)	
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)	
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)	
DE	MAPB	MAPB PROTEIN.	
GN	MAPB		
OS	Pasteurella multocida.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Pasteurella.		
OX	NCBI_TaxID=747;		
RN	[1]	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PM70;		
RA	Henderson I.R., Nataro J.P., Cappello R., Stein C.;		
RT	"Evolutionary origins of the autotransporter proteins."		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ277636; CAC14203.1;		
SQ	SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;		
Query Match	9.3%;	Score 966.5;	DB 2; Length 2712;
Best Local Similarity	22.0%;	Pred. No. 1.6e-32;	
Matches	629;	Conservative 338;	Mismatches 932; Indels 959; Gaps 128;
Qy	1	MNHLYKVIENKATGTFMAVAEYAKSH-----	26
Db	1	MNKYRVINSHVINTFTAVSELATSKGVKFSAISNPQELNSSIPATFKLSAIALVS	60
Qy	27	-----STGGSCATQGVSVRTLSE-----ARIAALAVLVIG---	58
Db	61	ILAFAPSQVLAQDVNAVNLTVSGTSTFTGAATFNNTATFNMMVTAGNVATTTLTVGKNV	120
Qy	59	-----ATLNGSAYAQOI---TTKIEIGOTNKINNT-----	85
Db	121	ATKDDVTTLTTTVNQKAAQSEVDNLKNTKADKSDVDNLKNTKAAQSEVDNLKNTKADKSD	180
Qy	86	-----LKG-----	89
Db	181	VDNLKNIKADKSEVNELNKSKADKSELTKLSNLVSLGVPEGTTTYTGKIFRVKSTLDD	240
Qy	90	ALATGEASIAFGSLSKAQGSQAIAIG-----SVKPDPNNGSNGNVS	131
Db	241	AVADQDSVAIGPRAKTEGVDAVALGHDSTANATESIAIGKAYAVK--TANKGI--AIGE	297
Qy	132	HAK-----GNESIAIGDVLAEGDASIAIGSDDLPLPKLNDL	168
Db	298	HARVGSKQDGVYVYDGCASSYVLGPKDGESSVAIGDKAVSRGEASIAIGK--AITSNKDA	355
Qy	169	KNEFKH-----LTHGHEI---LKKIQSTDGKIKY-----RRT	198
Db	356	KSOIAKNNTALGTNAQAIASDNSIALGNNAARTNKDSISIAIGDSAETKAAHSIAVGTTS	415
Qy	199	RAQGHASTAVGAMSTAOGHFSNAGFYATAEAYSLAVGLAAQATKQSSIAVGSNAKANA	258
Db	416	KALAAALALGKLAIEAKGTSSVAMGNTSKADGNSVAVGONTQSOTLOSNTIAIGSSAIANP	475
Qy	259	FAATATGNN-----TVNVLGR-----GVALGFGSQILDRDN---	289
Db	476	ERTISGLNAGKQEQADATGTRKHSQINIGENSGEVIGQLNIGIGAHAGTNNVYGHNTIAL	535
Qy	290	-----NTD---ASAYVPLGKTLADQY-----KATROGDST-----	316

Db 536 GTYACTNLNSETSGANVSIGHE-ANKVDOLITAVOKSTVVGQATKAASRSTALGAEATA 594
QY 317 ---DIFSIG-----NSNN----- 326
Db 595 LGLDAVAVGITSKAEQDKSVAIGANSTADSNVALGATSRAVAKEGGYLTGKQSSILVVS 654
QY 327 -----NNSIRPKLIINVGASROTDVAVNAQLKVLBEL----- 359
Db 655 VGHQDGAADHILRLRVLVADGVDEQDAATVAQLKVKTEKVTSDLOAQFNALSHAPTASE 714
QY 360 -----ANKRITFK-----GDGDNNSNSVERGIGLNTLTIKGDAOTNALTE 398
Db 715 IKYDTVPPSPPTGAENKIITLAKTRISNVAPAEALDADVNLGOVNIHVTKNAHYFSVND 774
QY 399 ANIGVVVT---DGNGLVKLAKELTGLTSVATNKITVSTNNNN-----AE 441
Db 775 TGINPVPHNDGATAKALAMAIAGONAAKAERSVAIGNNTTVNGESIGLGTYYKGSAR 834
QY 442 LQSGGLTFSPITGCTKTDKTV-YSID-CLAFETDNSNIATKGTTRITKKGIGFACPDNDGVD 499
Db 835 LODGVPRESETTVTKPSKNVKGIALGAGTITDGNNSIAIGSLAATSDK-----NPGAN 888
QY 500 ESKPYLDNEKLVGNSTLNGSLTVNNTTGNKIOVGANGIKFATVANNVANTSATVGTGA 559
Db 889 VDR-----AIAIGYNAVSAEKA--NAIGDRAVANSVKGNAFGSQAALGSAESTAIGTE 940
QY 560 RITEEKIGFA-----GTNDGVDEQ-----APYLDKE-RLK 588
Db 941 SKSEGOVAYALGTSKSHAKGLNSTAFCTNQVVSQNSGSIYAGELGNAKATVINGEGTYS 1000
QY 589 VGRVEIT---TDSGINAGNHKITGLTN---GIANT-----DAVTIKOLKD 627
Db 1001 LGNTSTLTANESGIFGNSNEIKAKENARIVGNKNTIGAIEEKHPVHPGTPPAAPVNDLKD 1060
QY 628 AKPTLNAGDGISNNSNGDLVDSG-----NIT-TPTYNISVKTTLKNSNGTSGNKK 678
Db 1061 IYVT---GYDNKISSDKLAKULSGLVFYVGHGNNTIAQLDPDPEETFTLTDSSVIGANNT 1117
QY 679 FVSNAHDNNSLVTARDLADYL-NKYNETADSALPCKFQVQNSNNAITVGRDITNGKTF 737
Db 1118 L---NTKGRNYFVLGNVNTATLNSVYLGLADSAYTTGNSTSSNNSYADMANGLNKSQYTF 1174
QY 738 NTLKLGNGVNTTNRAGTVTFGIDQSNGLTTPKLTVGSDFNGNRL---VIEQVPSADG 795
Db 1175 -----AGSOPVGVTVYGVGKERRVQNVASGLVTEAST--DAINGSQLFALRPLRFAGD 1227
QY 796 NST---KNIIKGLSPPLPSTASPSGRNIALG-----NTIEBKDKSNAASIDDLNAGF 845
Db 1228 NSTLSNPCKPGDVTVISSSSNOGKVVYGGENDGNKLTITADKNIGVTIANGDHTLEVR 1287
QY 846 -----NLKN-----NGKDKDFVSTYDVTDFIDGNATTATVT---YDEANQTSKVAYDVN- 891
Db 1288 AKTLSNLKDATFTGTDKTTINK-DCGMTITNG-ANTVSLTEGGLNNGKNKITNVAAGQNE 1345
QY 892 ---VDEKTE-----LTGDNQ---KQOLGVKTKTKLTSTNGNATFTSTDDHALVK 937
Db 1346 TDAVNYKQLNDLKAEGFGLTGEDGQTVKQALGT-AIKV--TGDDNKTIVTIDADGS--K 1400
QY 938 ASDIA-GNLTALAEHTTKGTANTALQFTVKKYVDENDKA--DDTNATITVGKDTSG-- 992
Db 1401 KLEIGLENQVTLGGEAKKNPAADGKL---TLKNQAGTDKVVLDGANG-TVGLTGADGAG 1456
QY 993 KVNTLKLKGNGLD-----IKTRDKGTVTTEG-----INTOSGLKAG 1028
Db 1457 AVITVK-KGRPTLDNAETPRIAYGNEEVATLNDG-LKFCANAGDVHNAKLNTQVDVKA 1514
QY 1029 DSTLLANN---GLSIKNTASNEQIOVCADGVKPFAMVNGVVGAGI-----DGTTRIR 1078
Db 1515 TANTVWDFDKGQNIIMTRVEGNTITVA-----LAKALSGLTSATFGDPASNPKDSTVINK 1569
QY 1079 DETGFTGTNCSLDKSKPHLSKODGINAGGKKITNIQSGELAKNS-----HDAV--TGGKI 1130

Db 1570 D--GLTITQGDNTVS---LTDDGLDNGNKQIKNVASGLTNTNGTATTSLDDAVOTNGVNV 1624
QY 1131 YDLKTELENKISSAKTAQNSLHEFSVADQGNFTVSNPYSDYTSKTSDDVITFAGENG 1190
Db 1625 GDLKTAINNITNGT---NPLGGFGLKDKAGNTF-----KQNLGETAQI--TGDSN 1669
QY 1191 ITPTKVNGKVVVRCIDQTKGLTTPKLVVGNNGKGVINSONGONTITGLSNTLANVTNDK 1250
Db 1670 VNTKV-----VD-----GONGKALEVSL---ANQLTLGRGPEANVPNAT 1706
QY 1251 GSVRTTEOGNIIKDEDKTRAASITVDVLSAGFNLQGNCEAVDFVSTYDVNFANGNTTAK 1310
Db 1707 G-----EAGKITLKDDKGTDRVVVDGSEGAISLTGPATOGAAAATAKIKVAECNPDLN 1761
QY 1311 VTYD---DTSKTSKVYVDV-NVDDTTI-----EVDKDKLGVK--- 1343
Db 1762 TSDDPANPNQNKTRITYDIAGPNGTITVTEQLATLNDGLKFGANTGVHDKALNTRVDVK 1821
QY 1344 ---TTLTSTGTGANKFALSNOATGDALYKASDIVAHLNTLSGDIQTAKASOAN--- 1395
Db 1822 GKAENTNWANFDAGON---IMTQISGNTIT---VALAKALAG-LDSATFGNPADGSKD 1872
QY 1396 ---NSAGYVDADGNKVI-----YDSTDNKYQA-----KNDGTVDKTEV----- 1432
Db 1873 GAVINKDGLTITEGDKTVKLTKEGLDNGNQIINVSGLKKTGDSVVALKDAEGSVLTNG 1932
QY 1433 -----AKDLVAQAQTPDGTLAQMYKSVINKEQVVDANKQ- 1469
Db 1933 VNVGDLKNAIKDVTSATNGGFGGLKDKAGAEFKQDLGTTAQITGDKNINTKVIDVPNSDK 1992
QY 1470 -----GIN-EDNAFY----- 1478
Db 1993 ALBISLANDITLKGNGADGVDSGLVNGKDGASVVLNGKDGSIGLTGPRQDGSKGKSAT 2052
QY 1479 -----KGLEKASDNKTKNAVT-----VCDLNAVAQTPLTFAGDTG-TTAKKLGE 1523
Db 2053 ISVKDGKAGVDGDKGTTRIVYETKDATGKPVVEEATLNDGMKFGVNDGKEVTRKLE 2112
QY 1524 TLTIKG---QTDNKLTDNNIG-----VVAGTDG--FTV 1553
Db 2113 TLDIKGLDAATVADNAKVSNNLGVKTNAGETGLEIVMKERTFSLVYVNGKDGDAAV 2172
QY 1554 KLAKD-----LNLNSVNAAGTKIDEK---GISFVDANGQAKANTPVLISANGLDL 1600
Db 2173 KFAKDKDGMGSIAAVTD-NDGNATGLTIKDKDGNPGVTF---NNDGR----- 2215
QY 1601 GGRVINSVNGKGTDDAANYQOLNE-----VRNLLGLGNONADGNQVNIADIKK 1649
Db 2216 ---ITNVTAGVDKDAVNVSQLDKGLAKATTKVEAGKNMTVPTVNVQDGSSTYTVVATED 2271
QY 1650 DPNSSSSNRPTVIKAGTVLGGKGNNDTEK-----LATGGVQVGVDK- 1690
Db 2272 NVNF-----TTVTITGNTVM---NDGVKVGDNVALTNEGLKAGDVTVTTAGINAGNKV 2322
QY 1691 DGNANGDLS-----NVWVKTKQDKGSKKALLATYNAACQTVNVTNNPAEADIRINEQG 1742
Db 2323 TGVAADGISPNSTDAVNGSQLNAVKETAEGWHLTANGADSSNV---KPRNTVDLNNITDG 2379
QY 1743 -IRFFHVNDGNQBPVVGVRNGRIDSSAGSKHSAIGFOAK--ADGEAAVATIGRQTA-GNQ 1798
Db 2380 NIVISKNTADKINVT---FGLADNINVKDSVVVGPKGANGKPGEGAVVINAEDGANGKD 2436
QY 1799 SIAIGDNAQATGDSIAI-----GTGNV-----VAGKHSQA-IGDPSTYKADNSYSVGN 1846
Db 2437 GISI---VGKDKRDAVAISGDKGVGTIGLTGPAGADGKNANAIIGVNDSVKGLDNGDKD 2493
QY 1847 NNOFTDATQTDVFGVGNNTITVTESNSVALGSNSAISAGTHAGTOAK-KSDGTAGTTTAT 1905
Db 2494 GNSKTRIVTKPNEEEQV-ATMNDGLVFGADGTEHAKLGTITVKVKGGDKNIETEVAG 2552
QY 1906 ATGTVKGFAGQTAAGVAVSGAS-----GAERRION-----VAAGEVSATSDAVNGSQLY 1955
Db 2553 DTLRVR-LKDNIDVKGINVTENLTVEKAKINMGNVNDGVADGEVNAATSKOAVNGSQLH 2611

QY	1956	KATOSI---ANATNELDRIHONENKANAGISSAMAMASPOAYIPGRSMVTGGIATHNG	2012
Db	2612	KVOOVNQNQATANKLGDHINKVDKDLRAGLAGATAVAFLOPNEAGKSIVSLGVGSYRS	2671
QY	2013	OGAVAGLSKLSRDGOWFKINGSADTQGHV--GAAVG	2048
Db	2672	ESIAVGVYARNSDNNKISIKLGGGMSRGDVNFGGSIG	2709
RESULT	4		
Q9F3X6			
ID	Q9F3X6	PRELIMINARY; PRT; 1299 AA.	
AC	Q9F3X6;		
DT	01-MAR-2001	(Tremblrel. 16, Created)	
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)	
DT	01-MAR-2001	(Tremblrel. 16, Last annotation update)	
DE	MAPA PROTEIN.		
GN	MAPA.		
OS	Pasteurella multocida.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Pasteurella.		
OX	NCBI_TaxID=747;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PM70;		
RA	Henderson I.R., Nataro J.P., Cappello R., Stein C.;		
RT	"Evolutionary origins of the autotransporter proteins.;"		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ277635; CAC14202.1;		
DR	SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;		
QY	1	MNHYKVFENKATGTFMAVEYAKSHSTGGSCATQ---VGSVRTLSFARIAALVLI	57
Db	1	MNKYRFLMAAQSVVWVSELAKA---GGKSASCKSALVNSVGSFSTLIAASVVL--	54
QY	58	GATLNGSAYAAQIITKIEIGQTNKINTLKGDLATGEASIAFGSLSKAQSGAIAIGSV	117
Db	55	---GS---GOVNAEVT-----	65
QY	118	KPDPNNGSNGVSHAKGNESIAIGGDVLAEGDASIAIGSDDLILPKNLDLKNFPHKLIH	177
Db	66	---GNTG---VSGDDKYCFYNASSQS---VIC	88
QY	178	GHEILKIQTSTDQKIKYRRTTRAGHASTAVGAMSYAQGHFSNAGFYATATAEAYSILAVG	237
Db	89	GDAATK---TTD---KTDNKPASVVIJ-----FG---ATNDGETNVAIG	124
QY	238	LAQAOTKOSSTAVGSKANAFATAIGGTVTVNLGRGVALFGSQILDRDNNNTDASAYV	297
Db	125	AKSKSKAASIAIGNAKALNOAIAIQNATANSDDWDISIG-----	166
QY	298	PLGKTLADQYKATRO--GDSTDIFSIGNNNNSIRKIIINVAGSRDIDAVNVAOLKL	355
Db	167	---RQGAEEQTEVSAEGRN---IAIGDGA-----	191
QY	356	VEELANRKTFFKGDGDNNSNVERGLNLTITIKGDAQTNALTEANIGVWTDGNGLKVLA	415
Db	192	---RKGVNNIAILGTAGDRLAGTHVLMGTVYNDAEAV-----	228
QY	416	KELTGLTSVSATNKITVSNNTNNNAELOSGLLTFSPITGTKTDTKTVSIDLKNTDNSNS	475
Db	229	--RSALTAGSTTKKINAKETDNK-----YYIEA-----SNT	258
QY	476	IATKGTTRTKKKIGFAGTNGDVDESKPYLDNEKLVGNSTLNSGSLTV-----NNTTGN	530
Db	259	VAL--GTRALATQLAAYVA-----IGQAKAFGNQSVAVNGTKASQGTATATGNSAHATGS	312
QY	531	KQIOVGANGIKFATVANNVANTSATVGTARITEEKIGFAGTNGDGVDEQAPYDLKERLKV	590
Db	313	SSIAIG-----GIVNGN--DTARTLASDNFTIAM-----GLSAQATKSD--AIAVG	354
QY	591	R---VEITTOS--GINAGNHKITGLTNGIAN-----TDVATIKQLKDAKPTLNAGDIS	639
Db	355	RNAKAAGINTVIGSYAGVTKTADAEOADPNKLEPATDAVFIGNKAGYKSNQRMQ--VS	413
QY	640	INSNNGDLVDSSGNITPTTYNISVKTTKLNSNGTSGNNKFSVNAHDN--NSLVATAKL--	696
Db	414	LKDSGEGVGVTEVNTIG--NSAKNTKGTNTVAISSRACQNVGECHONFAALIEAGONIK	471
QY	697	ADYLNKVNETAADSLPSFKYONGDNNNAITVGTDKTNGKTFNTLKLKGNGVNTTNRA	755
Db	472	GSDNIAIGHAGRSADPNTKL---NINNTISLKE-----SVSLA-----	508
QY	756	TGTVTFGIDQSNGLTTPKLTVGSDTNGNRLVIEQVPSADGNSTKNIKLSPTLPSTASP	815
Db	509	---NFGIAQGNKAKTDGLA-----SIAIGRNAEAVGGTANIAIGDSAS---ADA	552
QY	816	SGRNIALGNTIEEKD-----KSNAASIDVDVLNAGFNKLNKNGKDKDFYST	859
Db	553	SGA--IVLGTAKAOKSLTVDGKKYGAYSAIVIGTEAKATAQAAPAGKN--ENPKDAIAGT	609
QY	860	YDVFIDGNATATVTYDEANQTSKVAYDVNVVDEKTIETLTDGNGKKQLGVKTIKL--TET	918
Db	610	KABAHY-----ASTIALGFGAKSDTKAQAVSIGYN-----SNAK---GYQAIAGFSEA	654
QY	919	STNGNATTFSTDDDDHALVKASDIAGNLNTLAEIHTTKTANTALOTFTVKKVVDENDKAD	978
Db	655	KTENA-----GSSIA--FGTKAQ---TRASASIAIGMGAETGDFDGOALD	695
QY	979	DTNATVKGDKTSGKYNVTLKLGKNGLDIKTDKDGTVTFGINTQSGLKAGDSTTLNNNGL	1038
Db	696	GSDAVALGREA-----KAKRONAL-----AFGYKAVADHK-----	725
QY	1039	SIKTASNEIOVQADGVKPFAMVNGVVGAGIDGTTTRIDREIGTFGTNGSLDKSKPHLS	1098
Db	726	---DAVALGA-----GAETAAEETNEATVNEEFKYSGFAG-----IKPIAT	763
QY	1099	KDGINAGCK---KITNTOSGEIAKNSHDAVTKGKIVDLKTELENKISSTAKTAQNSLHE	1154
Db	764	---VSVGKDAERTITVNAGRIDKSTDAINGSOLY-----LALNAL--	803
QY	1155	FSVADEOENNFVSNPYSSYDTSKTSKTSKTSKTSKTSKTSKTSKTSKTSKTSKTSKTSK	1214
Db	804	---AG-----	829
QY	1215	LTVGNNGKGVINSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEDKTRAASIV	1274
Db	804	---GN-----VGNTL--VTNVLGG-----DAAIVKEGDE-----	827
QY	1275	DVLSAGFNLQNGEAVDEFVSTVTVNFANGNTTAKVTYDDTSKTSKVVYDVNVYDDTTIE	1334
Db	828	---AG-----	842
QY	1335	VKDKKLVGVTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQA	1394
Db	830	---TLTWSNIGG-----TGKG-----	842
QY	1395	NNSAGYVDADGNKVIYDSTDNKKYQAKNDGTVDKTKKVAOKLVAAQATPDGTLAQMNVK	1454
Db	843	---TIHDA-----	849
QY	1455	SVINKDQVNDANKKQGINEDNAFVGLKAAASDNKTNAAVTVGDLNAVAOPLTFAGDT	1514
Db	850	AVNNTAKASKTTVKEG---DNITV--TEEAAD-----	877
QY	1515	GTTAKKLGETLTIKGGQTDNTKLTDDNNIGVWAGTGTGFTVKLAKDLTNLSNVAGTKKIDE	1574
Db	878	---GSRITYTVATKKOV--KFDVSVVAGGTCKIDA	904
QY	1575	KGISFVDANGQAKANTPVLSSANGLDGGKVISNVGKG--TKDT--DAANVQOL----NEVR	1627

Db	905	NGLTFVDDQGTIDNTPPSLKSGTIDAGNOKVNTNVQNGNIADKSDKAVNGGOLFQAQGEV	964
Qy	1628	NLLGLGNDADGNQVNIADIKKDPNPGSSSNTRIKACTVLLGKGNNDTEKLATGGVQVG	1687
Db	965	NIIG-----GDTYNPETGEVAN-----TNIGGTGASTI-----	993
Qy	1688	VKDGNGANDLSNVVWVKQDGSKKALLATYNAAGQNTYVNNPAEIDRINEGGIRFFH	1747
Db	994	-----DEAKAVNTTAKAA-----	1007
Qy	1748	VNDGNQEPVVOGRNGIDSSAKHSHVAIGFOAKDGEAANAIGAIPQTOAGNQSTAIGDNAQ	1807
Db	1008	-----KTEVVOGENIVVTSAPGAN-----GFTVTVATAKE	1038
Qy	1808	ATGDQSTAGTCTGVNAGYHSGAIGDPS--TVKADNSVSVGN--NNQFTDADTDVPCV--	1861
Db	1039	VTEDKTT--VGSVYTDKNTDITGLSNKTLGGDNFAKNRAASEEQLNATQINLATILG	1095
Qy	1862	GN-----NITVTENSVALGS-NSAISAGTHAGTQAKK-----SDGTAGTTTATAGATGT	1909
Db	1096	GNAQNTNGNVAMVTDIGGCTCKNNINDAIKASRNEVKQGNVWVPTTCANGQTIYEVATAD	1155
Qy	1910	VKGFAQTAVGAVSVGASCAERRIQNVAAAGEVSATSTDAVNGSOLYKATQSIANA-----TN	1966
Db	1156	KVAF-DEVKVGGITDAT--TNKISGIAKGISDENSESTDAVNGSOLYELQOKIAKSGDNYN	1212
Qy	1967	ELDHRTIHQENKANAGISSAMAMSPQAYIPGRSMVTGGTATUNGOGAVAVGLSKLSDN	2026
Db	1213	ILNRRINKYDKDLRAGIAGANAAAGLPQAYIPGKSMWAVAAAGTYKGNALTAGMSRISDN	1272
Qy	2027	GQWVFKINGSADTQGHVGAAGVAGFHF	2053
Db	1273	GKVIILKLTGNTNSRGDFGASIGAGYQW	1299
RESULT	5		
Q9PC04			
ID	Q9PC04	PRELIMINARY;	PRT; 1190 AA.
AC	G9PC04;		
DT	01-OCT-2000 (TReMBLrel. 15, Created)		
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DE	01-MAR-2001 (TReMBLrel. 16, Last annotation update)		
DE	SURFACE PROTEIN.		
OS	XFL1981.		
GN	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		
OC	Xylella.		
OX	NCBI_TaxID=2371;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=9A5C;		
RC	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,		
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,		
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,		
RA	Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,		
RA	Kriegler J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.		

DR EMBL: U38617; AAC43721.1; --
SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;
Query Match 6.7%; Score 690; DB 2; Length 1098;
Best Local Similarity 24.7%; Pred. No. 2.le-21;
Matches 314; Conservative 151; Mismatches 439; Indels 366; Gaps 48;
QY 953 HTTKGTANTTA---LOTETVKKVDENDKADDTNATITVGKDGTSKGYNT-----996
DB 26 HTKCAATVAVAVLATLLSATVEANNPTVTKLAKAYGDANFNFTNNSIADAERKQVQEAY 85
QY 997 ---LKLKGNGLDITKDKGTVTGINTQSGLKAGDSTLLNNG--LSIKKTASNEQIQV 1051
DB 86 KGLLNLNEKNASDKLLVEDNT-----AATVGNLRLKGLWLVLSKNGTRNEKQ- 132
QY 1052 GADCVKFAVMNVCVAGIDGTRITRTRDEIGFTGTNGSLDKSPHLSDGKNAGKKTIN 1111
DB 133 ---OVKHA-----DEVLEFEG-----KGVQVTS 152
QY 1112 IQSGEIAKNSHDAVTCGKIYDLKTELENKISSTAKTAQNSLH--EFSVADEOGNFTVSN 1169
DB 153 -----TSNGKHTITFALAKDLG-----170
QY 1170 PYSYDTSKSDVITFAG--ENGITT--KVNKGVRVVGIDQTKGLTTPKLTVGNNGKI 1225
DB 171 ---VKTATVSDTLTIGGGAAGATTTPKVV-----TSTDGLKFAKDAAGANGDTTV 220
QY 1226 VINSQONTITGLSNTLANVTNDKGSVRTTEOGNIIKBDK--TRAASIVDVLSAGFNL 1283
DB 221 HLN-----GIGSTLDTL--VGSPTHIDGG---DQSTHYTRAASIKDVLNAGWNI 266
QY 1284 Q-----GNGEAVDFVSTYDTVNFANGNTTAKVTYDDTSKTSKVVDVNVDDTTIEV 1335
DB 267 KGVKAGSTTQSENVDVSVHYDTVEFLSADTETTTVTVDSKENGKRETKIGAKTSVKE 326
QY 1336 KDKKLGV-KITTLTSTGTGANKFALSNOATGDALVSRASDIAHLNLTSGDIQTAKASQA 1394
DB 327 KDKLFTGKANKETNKVDGAN--ATEDADEGKGLVTAKVIDAVNKTKGRIKTTDANGON 384
QY 1395 NNSAGYVD-----ADGN--KVIYDSTYDNKYQAKNDGTVDKTKEVAKDLVAQATPD 1445
DB 385 GDEATVASGTVTFASNGTATTATVNTGTDG--ITVKYDAKVGDLKLDGDKIAA----D 437
QY 1446 GTLAQMNVKSVINKEQVNDANKKOGINEDNAFVKGLEKAASDNK---TKNAAVTVGDINA 1502
DB 438 TTALTIVN-----DGKANNPKGVADVASTDEKKTAKGLVTA--LNS 479
QY 1503 VAQTPLTTFAGDTGT-----TAKKLGETLTIKGQDITNKLTDNNIGVVGAGTGTVKL 1555
DB 480 LSWTTTAAEADGGTLDGNASEQEVKAGDKVTFKAGK-----NLKVKQEGANFTYSL 530
QY 1556 AKDLTNLNSV-----NAGGTKIDEKGISFVDANGQA--KANTPVLNGLDLGKGVISN 1607
DB 531 QDALTGLTSITLGTGNNGAKEINKDGLTITPANGAGANNANTISVTKDGISAGSQSVKN 590
QY 1608 VGKGTKDTDAANVOQLNEVRNLLGLGNADGNQVNIADIKDPNCSGSSNRTVIKAGTV 1667
DB 591 VVSGLKFKFGDANFDPLTSSADNLTKQNDODAYKGLTNLDEKGTQKQTPPVVAQNTAATVD- 649
QY 1668 LGKGNNDTEKLATGQVGVQVDKNGDNLNVMVVKTKDGSKKALLATYNAAGQTNV 1727
DB 650 LRLGLWVTSADKTTGSGTEYHDOVRNNE-----VKF-KSGN-----GINVSGKT--- 693
QY 1728 TNNPABEADIRNEQGIIRFFHVNDG----NORPVQGRNGIDSSASGKHSVAIG--FOAKA 1781
DB 694 -----VNGRREITFELAKGEVVKNSNEFTVKETNGKETS-----VKVGDYKYSKE 738
QY 1782 D-----GEAAVATGROTA-----GNOSTAIGNDAQATQDSIAIGTGNVNAKHSAGIG 1831
DB 739 DIDLTGTQPKLKDNTVAAKYQDKGVKVSVDTEATITNK---GSGVYTCNQVADAI- 794
QY 1832 DPSTVKADNSYSYGVNNQ-----FTDATQTDVFGVGNNTVITESNVALGNSAISAGT 1885

DB 608 SNVYPGSEVDLKNSDGNLLITKTDS-----NDVTENLATALKVDLSLTGTNTAMTDTGV 661
QY 1448 LAQMNVKSVINKEQVNDANK--KOGINEDNAFVKGLEKAASDNKTKNAAVTVGDNLNAVAQ 1505
DB 662 TVGSMVTLGSTGLVTDGSPVTSSEGI-----SAGNQKITVNAAGTADTDAVNF 709
QY 1506 TPLTTFAGDTGTAKKL-----GETLITIKGQDITNKLTDNNIGV--VAGTDGTFVTKAKDL 1559
DB 710 SOLQAVSSTASKGNLNLASGANSNVAFGESVDLKNKDGNIVISKESGSDNVLFLNSSL 769
QY 1560 TNLNSVNAAGTKIDBKGTSF-----VDANGAKANTPVLNGLDLGKGLSVNNGKGTGD 1614
DB 770 -KLDKLTVDGDTVMTNGTVGSGVTLGSMGLVITDGPVTSVSSGKNAGSQKITVNAAGTAD 828
QY 1615 TDAANVOQLNEVRNLLGLGNDNADGNQVNIADIKDPNCSGSSNRTVIKAGTVLGGKNN 1674
DB 829 TDAVNLSQLN-----TAMAGSGAK 847
QY 1675 DTEKLTAT--GGVQGVQVDKNGDNLNVMVVKTKQDKGSKALLATYNAAGQTNVYTNPA 1732
DB 848 SVHYSTYDGGTQ-----GCNTYNGD-----867
QY 1733 EADIRNEQGIIRFFHVNDGQEPVVOGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQ 1792
DB 868 -----GATGTRSIAGVGVGTLASAEGATAVGS 894
QY 1793 TQA-GNQSTAIGNDAQATQDSIAIGT-----NVVAGKHS-----AIGDPS 1834
DB 895 AASGKGSTAGRNAVASADGVALGDGAKDARGAESYTGKYSGLQNNVTGVSVDGAS 954
QY 1835 TVRADNSYSYGVNNQFDTAT--QTDVFG-----VGNNI-----TVTESNSVALGNS 1878
DB 955 KGETRTVSNVADAKEATDAVNLQDLDRVQAQDANRYVDNKIESLSEGOFPVKVNSL--NN 1011
QY 1879 SA--ISAGTHA-----GTOAKKSDGTA-GTTTTAGATGTVKGFAGQATVG-----A 1921
DB 1012 SATPIAGVDATAIGVGATAGSADGSIAMGNKASASADNAV-----AIGNHSVADRAWT 1064
QY 1922 VSGCASAERRIQVNAAGEVSATSDAVNGSOLYKATQSTANATNELDHRHONENKANA 1981
DB 1065 VSGSAGSESRQVTKVNAAG---TADTDVNVSQL---NOGLITAKQYTDGVVGLRDRDGTG 1118
QY 1982 GISSAMAMASPOAYIPGRSVGTGIATHNGQGAVALVGLSKLSDNGQWVFKINGSADTQ 2041
DB 1119 GVAALATANLPQAYIPGRGWTSGVSSYRGQSAIAVGVSSVESGRWVFKFSGSANTRS 1178
QY 2042 HVGAAVAGAGTFH 2053
DB 1179 QVGIGAGVGYQW 1190
RESULT 6
Q48152 ID Q48152 PRELIMINARY; PRT: 1098 AA.
AC Q48152;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE ADHESIN (HIA).
GN HIA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NONTYPEABLE STRAIN 11;
RX MEDLINE=96332658; PubMed=8730864;
RA Barenkamp S.J., St Geme J.W. III;
RT "Identification of a second family of high-molecular-weight adhesion
RT proteins expressed by non-typable Haemophilus influenzae."
RL Mol. Microbiol. 19:1215-1223(1996).


```
Db 795 -----AKSGFELGLADEADAKRAFDDKTRALSAGTTEIVNAHDKVRFANGLNKTVSAAT 848
Qy 1886 HAGTQAK-----KSDGTAGTT 1901
Db 849 VESTDANGKVTTFVKTDVLPQLTYNTDANGKKITKVVKDGQTKWYELNADGTADMT 908
Qy 1902 TT-----AGATGVKGFAGQTAGVAVSG-----A 1926
Db 909 KEVTLGNVDSGKKVVKDNDGKWYHAKADGTADTKTGEVSNKYSTDDEKHKVVSLLDPNDQS 968
Qy 1927 SGAERRIONVAAGRSVATSTDAVNGSOLY---KATOSIANATNELDHR IHONENKANAGI 1983
Db 969 KGKGVVDINVANGDISATSDAINGSQLYAVAKGVTLVACQVNNLEKGVKNVKGRADAGT 1028
Qy 1984 SSAMAMASMPQAYIPGRSMVTGGTIATNGGAVAVGLSKLSLDNGQWYFKINGSDATQGHV 2043
Db 1029 ASALAASOLPOATMPGKSMVAIGAGSSYOGONGLAIGVSRISDNGKVIIRLSGTTNSOGKT 1088
Qy 2044 GAAVGAGFHP 2053
Db 1089 GVAAGVGQW 1098

RESULT 7
Q9F2D8 PRELIMINARY; PRT; 1107 AA.
AC Q9F2D8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SAPP PROTEIN.
GN SAPP.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID:601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277623; CAC14217.1; -.
SQ SEQUENCE 1107 AA; 113223 MW; F5C7CA651F5D51AB CRC64;

Query Match 6.5%; Score 677; DB 2; Length 1107;
Best Local Similarity 23.4%; Pred No. 7.4e-21;
Matches 303; Conservative 196; Mismatches 444; Indels 352; Gaps 58;

Qy 903 NGKQKLGKVIKLTETSTNGNATFTSTDDHALVKASDIAGNLNLAEEIHTTKGTANTA 962
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 SGRKLAVALV-----GLSSIMVSAD--ALANAGNDTGSINQNTDIATNTTSINN 71

Qy 963 LQTFVVKVDEN--DKADDTNATVVGDKTSGKVTNLLKKGKGLDITKDKDGTVTGGIN 1020
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 SNSVTTLTDDALLWDAASGT--FSASRGSASKITNL-AAQTAAADSTDAVNGSQLEDTN 128

Qy 1021 TOSGLKAGDSTTLNNGLSIKNTASNQIOVGADGVKFAVMNNGVVGAGIDGTTTRTRDE 1080
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 EKVQNTADITTTNTNSIN-QNTT-----DIATNTTSINNLSNV-----FTLLYDDA 173

Qy 1081 IGFTGTNGSLDKSPHLSKBDGINAGGKKITNIQSGEIAKNSHDVATGKGIVDLKTELENK 1140
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 LLWDAASGTFESASR-----NCSASKITNLAAGTAAADSTDAVNGSQLEDTNEKVDQN 225

Qy 1141 ISSTAKTAQNSLHFEVSVADEQGNFTVSNPYSSYDTSKTSDVI-----TF-AGENGIT 1192
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 -TADITNTNSINQNTT--DIATNTTSINNLSNVTTLTDDALLWDAADSGTFSASRNGSA 282

Qy 1193 TKV-NKGVVRVGIQDTKGLTTPKLTVCNNGCKGIVINSONQNT--ITGLSNTLANVTND 1249
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 8

Q9F285

ID Q9F285

PRELIMINARY; PRT; 3705 AA.

AC Q9F285;

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

```
Db 283 SKITNLAAGTLAADSTDAV-----NSQLYETNOKVDQNTSAIADINTSTNLSSD 333
Qy 1250 KGSVRTTEQGNIIKDEDKTRAASIIVLSAGFNQNGEAVDFVSTYDVTAFANGNTTTA 1309
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 -----NLSWN-----ETTNSFSASHSSTTN 354
Qy 1310 KVTYDDTSKTSKVYVYVNVDDTTTIEVKDKKLGKVKYTTTLTSTGTGANKFALSQATGDALV 1369
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 KITNVAAGELSEESTA-VNGSQLFETNEKVDQNTTIDIAANTTN-----IT 399
Qy 1370 KASDIVAHNTLSGDIOTAKGASQANNSAGVVDADGNKYIYDSDNKKYQAKNBDGTVDKT 1429
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 QNSTAIENLNTSVSDIWT-----SITGLTD---NALLMDE--DTGAFSANHGSGTSKI 447
Qy 1430 KEVAKDKLVAQAQPPDGTLAQMNYKSVINEQVNDANKKQGINEDNFAVFKGLEKAAADNK 1489
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 448 TNVA-----AGALSEDST-----DAVNGSQLYETNQK-----VDQN 478
Qy 1490 TKNAAVTVGDLNVAQAQPLTFAGDTGTAKKLGTELTKGGQDTOTNKLTDNNIGVVA--G 1547
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 479 TS-----AIADIN---TSITNLGTDALSWDDDEGAFSASHGTSCTNKTITNVAAGEIASDS 530
Qy 1548 TDGF-----TVKLAKDLTNLNSVNAGTK---IDEK--GISFVDANGQAKANTPVLSA 1595
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 TDAINGSQLYETNMLISOYNESISQLAGDTSYITENGTVGVKTYRTND-----579
Qy 1596 NGLDLGGKVISNVGKTRKTDAAANVQOLNEVRNLLGLGNADGNQVNIADIKKDPNSGS 1655
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 580 NGLG--GQDAYATGNG--ATAVCYDAVASGAGCLALGONSSSIEGSA-----LGSGS 629
Qy 1656 SSNRTVKA-----GTVLGGKGNNDTEKLATGGVQGVQVDKQGNANGDLSNWNVKTQ 1706
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 630 TSNRAITTGIRETSATSDGVVI--GYNTTDRELLGALSGLT--DGESYRQITNV-----679
Qy 1707 KDGSKKALLAT---YNAAGQNTVNTNPAEIDREINQGFREFHVNDGNEPVPVQGRNG 1762
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 680 ADGSEADAVTVRQLQNAIGA---VTTTTPT-----KYTHANSTEE-----716
Qy 1763 IDSSASGKHSVAIGFOAKADGEAAVAIGRQTQA---GNQSIAGDNAQATGDQSIATGTG 1819
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 717 -DSLAVGTDLSLWAKAKTIVNADAGIGIGLNTLVMAIDAINGIAIGSNARANHANSIANGNG 775
Qy 1820 NVVAKGHSIGAIDPSTVKAD-----NSYSVGNNN---QFTD-----ATQTDVFGVG-----1862
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 776 SQTT---RGAQTDYATYNDMTDPQNSVGEFSVGSSEGGORQITNVAAGSADTDAVNVGQLKV 832
Qy 1863 -----NNITVTESN-----SVALGSNSATSAGT--HAGTQAKKSDGTA-GTT 1901
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 833 TDAQVSRNTQGITNLNTQVSNLDTRVNINENGIDIVITGTSKTKYFKTNTDQADANAQAD 892
Qy 1902 TTAGATGTVKGFAGQTAGV-----AVSVGASGAERRIQNVAA-----1938
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 893 SVAIGSGIAAAENSVALGTNSVADEANTVSGSSTQORRITNVAAGVNNNTDAVNVQALK 952
Qy 1939 -----GEVSA--TSTDVAVNGSOLYKATQSGIA 1962
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 953 ASEAGSVRYETNADGSVNYSVNLGDSGGTTRIGNVSAAVNNDTDAVNYAQLKRSVEAN 1012
Qy 1963 NATN---ELDHRTIQENENKANAGISSAMANASMPQAYIPGRSMVTGGIATHNGGAVAV 2018
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1013 TYTQCKGEMGSKIKGLIENKRMGGGSIASAMAWAGLPQATAPGANMTSIAGGTFNGESAVAI 1072
Qy 2019 GLSKLSLDNGQWVFKINGSADTQGHVGAAGVAGFHP 2053
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1073 GVSWSVESGGWVYKLGTSNSQGDYSAAGAGFQW 1107

RESULT 8
Q9F285 PRELIMINARY; PRT; 3705 AA.
AC Q9F285;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
```


Thu Sep 13 14:19:33 2001

01-MAR-2001 (TReMBLrel. 16, Last annotation update)

YAPH. PROTEIN.
Yersinia pestis.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
NCBI_TaxID=632;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ277631; CAC14227.1;
SQ SEQUENCE 3705 AA; 370756 MW; 714PDF16455968C9 CRC64;

Query Match 6.4%; Score 658.5; DB 2; Length 3705;
Best Local Similarity 21.8%; Pred. No. 2.3e-19;
Matches 523; Conservative 325; Mismatches 925; Indels 631; Gaps 118;

QY 1 MNHLYKVFENKATGTPMAVAEYAKSH----- 26
DB 1 MNTIFKVINNASLVNVVVVSELAKGRITKSSRNLSIEGVLKPKFQSMVSKLFRKNLLAL 60
QY 27 -----STGGGSCATQGVSVRTLSFARIAALA-----VLVIGA--TLNGSAVAQOIT 71
DB 61 SLGSIVFLSTGPFAADITVSTQAELS-----AALSNGTYDKILGADITLISLVNMTS 116
QY 72 TKIEIGOTNK-----INNTLK-GDALATGEASIAFGLSLKAQGSQAIAIGSVKPDNNNGSN 126
DB 117 NOVVIDGGFGLTVNNTNYGLVSSGGTTLTQNNKIDSANYYSVMVL-----NGAN 171
QY 127 GNV-----GSHAKNESIAIGGVVLAEGDASIAIGSDDLVLPKNLDL 168
DB 172 TAVNVIYNNIDFLGSSQLIYMGAYGAATNSIMTFGDIL--NDVVNDRAQGEVKNKLA 229
QY 169 KNEPHKLIHGEILKKT-----STDGKIKYRRTAQQH--ASTAVGAMSYA-- 214
DB 230 TGRFHVHTGSSVTSFVSTGCGANTSTMDPASGADVKTIDRTGTGDLTSTGVNAFYTA 289
QY 215 -----QGHFSNA-----FGYATAEAAYSYLAAGLAQATKQSSIAVGS--NA 254
DB 290 DGASPELLANQVFSGTTNRGLEIGSYNSID-GFGSGVKIVLQSRSDGSLIISNGIDNA 348
QY 255 KANAFAA--TAIG-GNTVNLGRGVAL-GFGSQILDRDNNNTDASAYVPLGKTLADQYKAT 310
DB 349 TTNAAGINNAGSDANVIYNLTGTSILKATWTGILATKNANAS-----DIY-IR 397
QY 311 RQGDSTDIFSGNSNNSSIRKIIIVGAGSRDIDAVNVAQLKVE-ELANRKITPKGD 369
DB 398 SAGDITAATGISAHTNGTGV--KIKNDGTTTSTAGIAISSAKISELSVDNTDGTITAT 455
QY 370 GDNNSVVERGIGNTL--TIKGAOTNALTANIGVVDGNGLKVAKELTGLTSVSAT 427
DB 456 AGTVNVLASALNLEFGGTINTSATANGITAG-----TEGGHTLTDLTINLLG-TGIALS 510
QY 428 NKITVNTNNNAELQSGGLTSPITGKTOKTVYSIDGLKFTDNDNSIATKGTTRTKK 487
DB 511 NVAGVNLTSN-----VTLNTNGT-----ALNSLTGLTLVDSLNG--RNTINIEGA 555
QY 488 KIGFAGTNDGVDESKPYLDNEKLK-----GNSPLNSGSLTVN----- 525
DB 556 GIGIAATNTLENT-----FPAEALDINVNGAGIGIATGGVNLASNLIIINVANTLGTAL 611
QY 526 -----NPT--GNKIQVANGCIKATVANNVANTSVTARITEKIGFAGTNDGVD 576
DB 612 QITDGDINTTTIGN-BIQLNAEN--ATAINFGLSSKTLNNNGTIKGSVIFAGVAD--- 664
QY 577 EQAPYLDKELKVRVEITDSCINAGNHKITGLTNGIANTDAVATIKOLKDAKPTLNAGD 636
DB 665 -----HIINNNTLDGT-----LITGAGN-DTLVLIDSSSQSNDVINLGD 702

QY 637 G-ISINNNGLDVS-----SGNITPTTYNIVSKTKKLNS-NGTSGNNKFSVSNADHNSL 690
DB 703 GNSVTIQNGATVSSIIITGNGNDFTINGMSVSGTSLGSLDAGTGLATXNXXAST--EL 760
QY 691 VTAKOLADYLN-----KVNETADALSFPKQVQDGNNSNAI-----TVGKDTN 733
DB 761 AAATSLQGFNTINLVDSHTLVDDDDNIGSGMV-NIDSSSELLFGSTFGILHATLACGTG 819
QY 734 GKTFNTLKLKGNGVNIITNAT-----GTVTFGIDQSNGLTTPKLTVGSDTN- 781
DB 820 SAIVN-----NSANVSLEQASMEFAGTWQVNOGGALT--ASNSNOLGSAGIKLGDGLNL 870
QY 782 GNRVIEQVPSADGNTKNIKGLSPTLPSIASPSGRN-----IALGNTIEEKDKSNAASI 837
DB 871 DNIALENHV--LTGNGTLNVAKNLATAFDFGSGFAGFSGVNLTKTTFALSADNAAAL 928
QY 838 DDVNLNAGFNKNGKDKDFVSTVDVD-FIDG-NATTATVYDEA---NQTSKV--AYDV 890
DB 929 -----ASATLK-----LSDDSVTTVGTDTDRTLHGLDLSGGTLIFDGAVPQSQTSVGVVTVDL 980
QY 891 NVDEKTIETLTDGNGKQKOLGVKTIKLTETSTNGNATTF--TDDHA-----LVKASDIAGN 944
DB 981 ALNSGTVNIITGSG-----SWDNTDPLATNVSTLEQDRAGSTLELINATNVGD 1028
QY 945 LNTLAEIEHTTKGTANT-----ALQTFVTKVKVDENDK----- 976
DB 1029 IDALDLVNGTALTSGTGQVQSAIQGGSTVANAIHNYGLASSNSNGDSLGVNVTLSAL 1088
QY 977 ---ADDTNATVKGKG--TSQKVNTLKLKNGKGLDKTK-----DGTVT--- 1016
DB 1089 ELLADGADALLATATESGLTANRVNLNABELFGVGLVVDQAQNGALTLANGSNRYEGTTVTA 1148
QY 1017 -----FGINTQSLKAGDSTTLNNGLSIKNTASNEQIQVGADGVKFA--MVNNG 1064
DB 1149 GELILGANGAFQGTSLDIIASGASANINGVSTQVATNVGTVTLGSGGVLTSGLLTNGG 1208
QY 1065 VV--GAGID---GTTTRITRDEIGFTGNGSLDKSPHLSKDGINAGGKKIINIOSGEI 1117
DB 1209 ILDLTGALNLTXXGASTVAG--GLTGA-GTLINGNLSVSAANSGLSGQTHI--ADV 1262
QY 1118 AKNSHDATVGGKIYDLKTELENKISSSTAKTAQNSLHEFSADEQGNFTVSNPYSSYDTS 1177
DB 1263 A-----SVT-----LTDYTGTSAVEVLGTLNLNGANAAMTVLSGDGTI 1303
QY 1178 KTSDVITFAGENGITTKYKGVVRVIGIDQTKGLTTPKLTGVNNGKG-----IVINSQNG 1232
DB 1304 NTNAAVTLSGNNSFS-----GAHQIGTD-----GELTVQASNLGASSATVNLGLTIS 1351
QY 1233 QNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEKTRAASIVDVLS-----AG 1280
DB 1352 HLILNGVSESIANVL--GVAGSTV--DIIGGADTALTANNSGFLGOYALAGNSKLTVAS 1407
QY 1281 FNLQNGEAVDFVSTVDVYNFAN-----GNTTT-----AKVTYDDTSTKSVYVD 1325
DB 1408 TNNLGAASSVALAGAGDITLSLGFNGTFCNSVTGSGVLQVTDDAEVTLTSSNGVSNV-- 1465
QY 1326 VNVDDTIEVKOKKLGKVTTLTS--TGANKFALSNOAT---GDALVKA--SDIVAH 1377
DB 1466 -----TIDIADATLNDLIDIALFNHVLGTNGLLNVAKNDASTAFDFGTVGAFSGIVNL 1519
QY 1378 LNT-----LSGDIQTAKGASQANNSAGYVDAGNKVYDSTONKYYQ 1419
DB 1520 TMTFALSADNAAALARATLKLSDSDSVTVGAT--DRTLHGLDLNGGTLIFDGSPP---Q 1574
QY 1420 AKNDGTVDKTEKVAOKLVA-----QAQTPDGTLAQNMNVKSVINKEOVNDANKQ 1469
DB 1575 SQANGVVTVTDALNSGCTISITGAGWNEHPVTPPNVSLLEQDRCDLILLELINAANYTG 1634
QY 1470 GINEDNAFVKGLEKAASDNKTKNAAVTVGD--LNAAVQTPLTTFAGDTGTAKKLTETLI 1527
DB 1635 NANNLDDLVDGTALTSTQGTQGVESAIOQGGSTVANAIHNYGLTSSNGGSGGLVYNYTLS- 1693
QY 1528 KGGQTDNKLTDN--NIGVVAGTGTGTVKLAKDLTNLNSVNGAGTKIDEGISFVDANGQ 1585

QY	932	DHALVKASDIAGNLTAEIHTTKGTANTALOTFTVKKVKVDENDKADDTNATVVGKDGTS	991	QY	1789	IGROTQAGNQSIAIGDQAATGDSIAITGTGNVAG	---	KHSGAIGDSTVKADNSYVG	1845
Db	1108	DOSLSAK--IKNIN-----IGTVAGGA-----IYTLDAINDN---FDLN-----	1144	Db	2083	SGYQSTNGGIIIGFDYNI-----DNSIVIGAAYTMADSKVKHKNDKNGDRTKAKSNISYI	---	---	2138
QY	992	GRV-----NTLKLKNGKGLDKTDKGVITFGI---NTQSG-----LKAGUSTLNNGLS	1039	QY	1846	-----NNQOFTDA-----	---	---	1882
Db	1145	GMVEKHODSILEKNSNTN---DRTITLTSALDPGNNOFGIILITDITNKLITIDNG--	1199	Db	2139	GLYNMLTNNFFVEAIGVVGGRNKKIKYKRIITITDQIAIGKFINFYFVELLGGVNYLIS	---	---	2198
QY	1040	IKNTR-----SNEQIQVGADGVKFEAMVNGV---VGAGIDGTTTRITRD-EIGFTGTNGS	1089	QY	1883	AGTHACT-----	---	---	1931
Db	1200	--NVATICTANMLK-----QUTFASIDNGALAKVGINVENVTLNKIDIELNEVNANVL	1253	Db	2199	---HRTIITPMFGMRVATFKNGYKNNNTTFQNLISIKKNYDKFETILGLNSV	---	---	2248
QY	1090	LOKSKPHLSKDGINA-----GGKKTINTQ-----SGETAKNSHDAVTC-----GKIYD	1132	QY	1932	RIONVAAGEVSATSDAVNGSQL-----YKATQSTANATNELDHRHONENKANAGISSA	---	---	1986
Db	1254	FNKNTYTTATGTINGHVDPOGNAGVINLMDIDIDGSTVTGDN---VNGTLNFGSGKVTG	1311	Db	2249	-----THYLSQDIIIIKPELHWFINTYCKNKLFPNIDARLD-----	---	---	2287
QY	1133	LKTELENKISSAKTQNSLHEFSVADQEG---NNFT-VSNPYSSYDTSKT-----	1179	QY	1987	MAMASMPQAVIPGRSMVTGGIATHN	2011	---	---
Db	1312	LINNIVMLQAGADVLSADGSAGSYTEICGNGNNLTFAANSHLTTDINKTGGQDLNVF	1371	Db	2288	LTITRFKPAKITYN---LGGGISTKN	2310	---	---
QY	1180	-----SDVITFAG-----	1207	RESULT	10				
Db	1372	INGSVSGSIGANAAGDIIINAGSVNFSNTLKSIGNIVISDQATMOVNNVTATDISGN	1431	ID	Q9VTK8	PRELIMINARY;	PRT;	2586	AA.
QY	1208	KGUTPKL-----IVGNNGKGIIVINSONGONTIT-----	1237	AC	Q9VTK8				
Db	1432	ANNGTLKLNHHPITINTSTLGNNAIG-TIEVANNDVTITGLQAOIHFSNATQAATLT	1490	DT	01-MAY-2000	(TREMBLrel. 13, Created)			
QY	1238	-GLSNTLANVTNDKGSVRTE-----QGN--IIFKBDKTRAASIVDVLSAGNLOGNG--	1287	DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)			
Db	1491	LGAASQVNTTATAGNHTLEVTDFDGTNDGIIGDAN-NRLKSI-----ELTGNCTIV	1541	DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)			
QY	1288	--EAVDFSVTYDVNFANGN---TTTAKVITYDTSKTSKVYDVNVDDTTI-EVKDKKL	1340	DE	CG18331	PROTEIN.			
Db	1542	TINSPHYVSSITTTANNAQGNVKLNIEGGITYDLGSKISLANVQISEDTITIGDVIYSKYL	1601	GN	CG18331				
QY	1341	GVKTTTSTGTGANKFALSQATGALV-----KASDIAVHLNLTSGD-IOTAKGASQ	1393	OS	Drosophila melanogaster (fruit fly).				
Db	1602	NIDAGKTINFDGNNPNRNLDPDALIDLVLPRLSLFNFYFTDIKADNLNFADDTAT	1661	OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
QY	1394	AN-NSAGVVDAD-CKNVIYOSTDNKY---YQAKNDGVDTKKEVAKDKLVAQAQTPDGTILA	1449	OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
Db	1662	ANFKDAVVIDAHIDNGILFNDNAWLTQETKANII-----ELASDKF-----LL	1708	OC	Ephydroidea; Drosophilidae; Drosophila.				
QY	1450	QMNYS-----VINKEOVNDANKKQGINEDNAFYKLEKASDNKTKNAAVTVG	1498	OX	NCBI_TaxID=7227;				
Db	1709	QNKIKATLIADNANVLDDNVVNTNLNVRDVLDA---NYELKVTGNVTHNGLITI-	1764	RN	SEQUENCE FROM N.A.				
QY	1499	DLNVAQPTPLFAGDTGTAKKLGTTLTIKGGOTDTHKLTNDNNIGVVAGTGDGFTVKLAKD	1558	RP	STRAIN-BERKELEY.				
Db	1765	---FDTALQGGHILVSQGSNVDMSDL-DNLI-----IKIAHSD	1803	RC	MEDLINE-20196006; PubMed=10731132;				
QY	1559	LTNLNS-----VNAAGTKTDEKIGISVDANGAKANTPVLSANGL-----	1598	RX	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
Db	1804	ITNITSYDKHQIVKLETGAIYTPQTKVVIDASEQNKVKKVADANGLVLLTDTGGRD	1863	RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
QY	1599	DLGGKVIS-----NVGKGTDT-DAANV---QOLNEVRNLLGLGNADGNQVNIADIK--	1649	RA	George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,				
Db	1864	DTGGRODTRGRGNTDNGCRDNCVGNISNSNEA-----GGSSDKN-YGITDVPVIF	1916	RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
QY	1650	DPN-----SGSS---NRTVIAKGTVLGGKGNNDTEKLATGGVGVGVDKDG--	1692	RA	Sutton G.G., Rogers Y.-H.C., Blazej R.G., Champagne C., Pfeiffer B.D.,				
Db	1917	DSPILDYTKNIVVAGIANQINHVKGFTNTDAGKLLNDLGEMSPNRTETLDRLSNR	1976	RA	Brandon R.C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
QY	1693	-NANG-----DLNSNVWK-----TQKD-GSKKALLA---TYNAAQGTNYVT	1728	RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
Db	1977	INVNGLNEGVSGLNGIEVENFLTDIAINMDFNTAKEIGNRLEELSDANTVGLNKTNTLL	2036	RA	Abriel J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,				
QY	1729	NWPAEAIIDRINEGIRFFHVNDGNPEVVGGRNGIDSSAGKSHSVAIGFOAKADGEAAVA	1788	RA	Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
Db	2037	NNK-----INLKLNTNNQAIIAAGDE---DNIVGTGWGNSFYFKIKQNSKNS-A	2082	RA	Besson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				

Thu Sep 13 14:19:33 2001

143 GDVLAEGDASTAIGSD--LYLPKNLDLKNFHEHLKIHGHEILKKTOTSTGDKIKRRTTR 199
 172 G-----ATAGTITQISATPALINLAGADGNDH-----GEITVNTRTSFTGIGTKTL 221
 200 AOGHASTAVGAMSQAQHSNA-FGTATATAEAYSLAVGLAAQATKOSIA-VGSNAKAN 257
 222 ---HAVT-----FNNAGAGVSTAGASATTVMIGEDAGNVASVVQGLITGAV 267
 258 AFAATAGTIVNLRGVALGFGSOILDRONNTDASAVPLGKTLADQYKATQGGSTD 317
 268 NFAAD--GALTANN-----GIAGAVTTANNNG-----TL-----TVGAGDVTG 304
 318 IFSIGNNNNNNSIRKIIINGAGSRDT-DAVVAQLKLVEELAN-----RKITFGK 368
 305 AICV-----NGCNILKQVLFNGASNVATIDATNVT---INNAANVTAGAITAVNFAA 356
 369 DGNNSNSVERGLNLTITKGDQOTNALTEANIGV-----VTGNGLVKVLAKELTG 420
 357 DGALTANN---GIAGAVTTANN-NTGTLT---VGAGDVTGAIGVNGGNILKQVLFNGASN 409
 421 LTVSATNKITVSTNTNNNAELOSGLTFSPITGTYKTKTVYSIDG-----LKF 469
 410 VATIDATN-VTI---NNAANVTAGAITAAYN-----FAADGALTANNNGIAGAVTT 457
 470 TNDNSIATKGTTRITTKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGLTVNNTTG 529
 458 ANNTGTTLTVGAGDVT---GAICVNGG-NILKQVLFNGASNV--ATIDATTVTINNVA 510
 530 NKQIQVGA-----NGIKFA-TVANNVANTSATVGTARTTEKIGPAGTN 572
 511 N-VTAAGAITAAVNFADGALTANNAGAVTTANNNGT-TLVGAGDVT---GAIGN 564
 573 DGVDQAPYLDKRLKVRVEITDGSINAGNHKITGLTNGIANTDAVTIKLQDAKPTL 632
 565 GG-----NILKQVLFNGASNVAT---IDATNTVINNAANVTAGAITAAVNFADGALT 615
 633 NAGDGI-----SINSNNGDLVDSNGNITPTTPTNISVTKTLKNSNGTSGNN--KFSVNAH 685
 616 TANGIICAVTTANNNTGTLTVGAGDVT-----GAIGVNGGNILKQVLFNGA 662
 686 DNNSLVTAOLADLYNKV--NETADSLP---SPKVGONSNNALITVGKDTNGTKFTNTL 740
 663 SNVATIDATTVT--INNVAANVTAGAITAAVNFADGALTANNNGI-- 707
 741 KIKGNGVNIITNTRATGTVTEGIDQSNGLTTPKLTIVGSDTNGNRLVIEQVPSADCNSTKN 800
 708 -----GAVTTANNNTGTLTVGAGDVTG-----AIG--VNG-----GNILKQ 741
 801 II-KGLSPTLPSIASPSGRNIALGNTIEKDKSNAA-----SIDDLVNLNAGFNKNGKD 853
 742 VLFNGAS-----NVATIDATNTVINNAANVTAGAITAAVNFADGALT-----NNG-- 790
 854 KDFVSTYDVTDFIDGNATTVTYDEANQTSKVAYDVNVDEKTLTGTGDNCKKQL----- 908
 791 -----IAGAVTTAN-----NNTGTLTVGAGDVTGAIGVNGGNILKQVLFNGA 832
 909 -GVYTKTLTETSTNGNATFTFDDHALVKASDIAGNLTALAEETHTTK-----GTANTAL 963
 833 SNVATIDATNTVINNAAN-----VTAAGAITAAVNFADGALTANNNGIAGAVTTA-- 883
 964 QTFVKKVDENDKADDTNATITVKGKGTSGKVNTLKLKNGKGLDIKTDKGTVTFPOINTQS 1023
 884 -----NNNTGTLTVGAGDVTGAI-----GVNGGNILKQ-----VLF--NGAS 918
 1024 GLKAGDSTLLNNGLSINKNTAS---NEQIQVAGDGVKFAVNVGVVVG-----AGIDGTTRI 1076
 919 NVATIDATTVTINNVAANVTAGAITAAVNFADGALT-----NNGIITGAVTAGGNGT--L 974
 1077 TRDEIGTGTNGSLDKSPHLSKDGINAGGKLTINIQSIEAKNSHDVATGKIKYDLKTE 1136
 975 TTAGVGVGTGAVGT-----NAASIKVLN---AKVDNAGAADLVFTSDIY----- 1013
 1137 LENKISSTAKTAQNSLHEFVSADQGNNTVSNPYSDYTSKTSIDVITTF-AGENGITTKV 1195

1845 IPNSGRKLNSSSTSTTV---TSSSSSKPQTKYSWSSSSKKNNGGKNKYWKRTKTK 1901
 1710 SKKALLATYNAAGTQVYNNPAPADIRINEQGRIRPHVVDNQOE-----PVV 1757
 1902 SRK-----NNGSSSTIVGE---ESSDSLTDAGVDVDTQGNGLNDEGSSOSTVTSLLPVV 1952
 1758 QGRNIDSASG-----KHSVAIGFOAKADGEAAVAIGROTQAGNQSIAIGDNAQAT 1809
 1953 DTSADVQNSSESLSTENTNTYTSKSKFVKPSNQSSISASKTKTVTSTSSPPNVKSS 2012
 1810 GQSIATGIGNVAVKSHGSAIGDPTVKADNSYSVGNNOPTDA-TOTDFVGNGNITVT 1868
 2013 SKK-----TSN--SKK---SVKTSSTITTTSSDPGSSSITGIPQNDIKSL-NQVTTT 2061
 1869 ENSVALGNSAISAGTHAQAKSDGTAGTUTTAGATGVKGF--AGQFAGVAVSVA 1926
 2062 TSSVSQGVFPSSPVVVKTKETSVSKDGKTRSSSTTTTTTTTITTKGNSQSGTLTLPADV-GL 2120
 1927 SGAERRIQVNAAG-----EVSATSTDAVNGSOLYKATQSTANATNELDRIHONE 1976
 2121 KSKTKTTTSTKTKLSDILSLPEVDASI--AVNGDESRASIKDNTILSKIDLSLPKLD 2178
 1977 N--KANAGISSAMAMAPQAIPIGRSM-----VTGGIATHNGOGAVAVGLSK---LSD 2025
 2179 ASLVNNGGKSSKSSSTTTTTTTSTKGNKVSLSLPEVDASIAV-NGDARSASIKDNTILSK 2237
 2026 NGQWVFKINGSADTQ 2041
 2238 IDLSLPKLDASLVNG 2253

RESULT 11
 O9XC47 PRELIMINARY; PRT; 2106 AA.
 AC O9XC47
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE OUTER MEMBRANE PROTEIN A.
 GN OMPA.
 OS Rickettsia australis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID-787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PHS;
 RX MEDLINE-20487299; PubMed-11034486;
 RA Stenos J., Walker D.H.;
 RT "The rickettsial outer-membrane protein A and B genes of Rickettsia
 RT australis, the most divergent rickettsia of the spotted fever group.";
 RL Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
 DR EMBL; AF149108; AAD39531.2;
 SQ SEQUENCE 2106 AA; 209075 MW; A12B28F54BEEDECC CRC64;

Query Match 5.4%; Score 557; DB 2; Length 2106;
 Best Local Similarity 21.9%; Pred. No. 2e-15;
 Matches 509; Conservative 247; Mismatches 790; Indels 774; Gaps 125;
 QY 1 MNHIYKVFINKATGTFMAVAEYAKS-----HSTGGGSCATQGVSVRTLSPFARIAAL-- 52
 DB 1 MANISPKLFOKAHKSLKALETTTTPAAILMSSSGAWGVAAGVI-SVNDAAFSNRAVANN 59
 QY 53 -AVLIVGATLNGS-----AY-----AQOITTKIEIGOT-----NKINT 85
 DB 60 WNETAGGAANGHADGPQDNEAFTYCGNHTITADEAGRIITAINVAGTTPVALNSTQNT 119
 QY 86 LKGDALATGE---ASTAFGSLSAQGSQAATGSKVPDPNNGSNGVSHAGNESIATG 142
 DB 120 SVGSIVTGGNLLPVTIADGKSLTGTGKAVA-----ADHGFDAADNIT-GLGAIALG 171


```
Db 1014 -----ARTV-----NF-----DIAGGAGVTTQVGNLIATNV 1042
Qy 1196 NKGVVVRGIDQTKGLTTPKLTVGNNGKGVINS-----QNGQNTITGLSNTL 1243
Db 1043 NFG-----GNAQGGTFLXUNGPKVGSYTLTGTTANGNATINI-NTL 1082
Qy 1244 ANVTNDKSVRTEQGTGNTIKDEKTRAASI-----VDVLISAGFNLOGNEBAVDFVSTYTV 1299
Db 1083 GLVTABNAXIGTVAQINI--QDNKTFAINKNADIEIL-----NAQIDFKGANSKL 1132
Qy 1300 NFANGNTTAKVTYDDTSKTSKVYVDNVDDTTIEVKDKKLGKVTTLTSTGTGANKFAL 1359
Db 1133 FLVNSAT-----DDRVVTIKN-----DLPATGATGGMMLL 1163
Qy 1360 SQAATGDALVKASDIVAHLNLTSSDIOTAKGASQANNSAGYVDADGN---KVIYDSTDNK 1416
Db 1164 FG-----TANLITLQGD-----GAXKALGTAGNKLASLNVLGKVAFNID-- 1203
Qy 1417 YQAKNDGTVDKTEKAVDKKLVAQAQTPD-GTFLAOMNVKSVINKQVNDANKKQGINEDN 1475
Db 1204 -----TTNLVAFNIITTFQVDVGGIT--NOISVIN-----IGAANASLGL---- 1242
Qy 1476 APVKGLEKASDNKTKNAAVTVGDLNVAQAUTPLTFAGDTGTTAKKLGELTLTIKGGOTDN 1535
Db 1243 -----PAAAGSYTIDAN-----GGWVGILAN--GQTINF--AHEDAE 1275
Qy 1536 KLTDNIGVAGTGTGTVKLAKDLTLN-----SVNAG-----GKIDBK 1575
Db 1276 LVLQNS--AAGNSTITLNAALDPCAPNKGKLAVDGAGGAGRVILASVRNATYGTAVNKL 1332
Qy 1576 GLSFVDANGQAKNPVLSANGLDLG-----GKVIS--NVGKGTDKTDAAVQOOLNE 1625
Db 1333 KLEFRGNTQFQIDIF-ANNLEFPVPAATYNNKINDXNLSFGAATALTQNGNI----- 1386
Qy 1626 VRNLLGLGNDNAD-GNOVNIADIKKDPNSSSNTVVKAGYVLGKGNNDTEKATGCV 1684
Db 1387 -----NCNVDFNQAAVITL-----GANKNIT--GSVTSNGVNGT-TIATGAS 1427
Qy 1685 QVGVDKDNANGDLSNVVWK-----TQDKGSKALL--ATYNAAGOT 1724
Db 1428 TI-----NGHI?NIAMLKVKAGAVSITQGGNTSITEIQNGTALLTLPANFNLTG-- 1477
Qy 1725 NVVTNNPAAIDRIEIQIRFHVNDGQEPVVGQNGRIDSSASKHSAVIGFAKADGE 1784
Db 1478 -----SINKTGOALKLNFNGGSRGVV-----GT 1503
Qy 1785 AAVAIGROTAQGNQSTAIQDQNAAT-----GDQSI A--IGTCNVVAGHSGAIGDPSTVK 1837
Db 1504 AANSVGDIITTTGTTNFPASSVNAKGAATLCCTTSPADTFNTGAVTLAKAS----- 1553
Qy 1838 ADNSYSVGNNOFTDATQTDVFGVGNNTITVTSNSVAL-----GNSAISAG-----THAG 1888
Db 1554 -----ITNFAKNVATSF--AANNATINFGNSLAENSTIGSCTTLTLCASQVITYG 1603
Qy 1889 T-----QAQSDG-----TAGTITTAGATG-----VKGFAQQTAV 1919
Db 1604 TGSFTDTLTLNTPFGAASDGHILKSGSTLDSGVSTLALVVTATNPDINNIPNTRY 1663
Qy 1920 GAVSVCASCAERRIOWAAGESATSTD-----AVNGSOLYKATOSTANATNELDHR- 1971
Db 1664 TVISAETAG--LKPTPAGNVKVTVNDRNRVNTFFDESTLTILFAEDIAEEVIEDFRP 1720
Qy 1972 -----IHQNEKANAGISSAMAMSPQAYIPGRSMV-----TGGIATHINGQAV 2016
Db 1721 GGPLANIPNAANV--KKSLELMQAPIGSDACLAFDNMGKLNPIQAAADVTHLMQDVV 1776
Qy 2017 AVGLSKLSDNGQWVK-----INGSADTQGHVGAAGVAG 2050
Db 1777 KPSDTIAAINNOVILSNISSSLINLARMADAIQAPAAVAG 1816
```

RESULT 12

```
Q9JY30
ID Q9JY30 PRELIMINARY; PRT: 2514 AA.
AC Q9JY30;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HEMAGGLUTININ/HEMOLYSIN-RELATED PROTEIN.
GN NMB1768.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback J.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT WC58."; 287:1809-1815(2000).
RL Science
DR EMBL; AE002526; AAF42109.1;
DR TIGR; NMB1768;
SQ SEQUENCE 2514 AA; 265615 MW; 95643A671B3BC268 CRC64;
```

Query Match 5.3%; Score 553; DB 2; Length 2514;

Best Local Similarity 20.0%; Pred. No. 3.8e-15;

Matches 503; Conservative 332; Mismatches 85; Indels 782; Gaps 121;

```
Qy 4 IYKVIENKATGTFMAVEYAKSHSTGGSCATQVGS--VRTLSFARIAALAVLVICATL 61
Db 5 LRVIPNRRKGAIVAAETTKRE---GKSCADSDSGSAHVKSVPFG----- 47
Qy 62 NGSAYAAQITGKTIEIGTNRKINNLTGDA--LATGEASIAFG-----SLSKAQGSQAI 112
Db 48 -----TTHAPVCRSNIFSFSLGFSICLAVGTANAFADGIADRAKAPKQATIL 98
Qy 113 AIGSVKPDN-----NGSNGNVGSHAKGNESIAIGGDV-----LAEG 149
Db 99 QTCNGIPQVNIQTPTSAGVSVNQYAOFDVGNRGAILLNRSRNTQTQLGGVIQGNPWLAR 158
Qy 150 DASIAIGSDDLLYPKMLDLKNEPHKLHGHEILKKTQTDGKIKYRTPAQGHASTAVG 209
Db 159 EARVVV-----NQINSHSSQMGNYIEVGGRAEVVIANPAG 195
Qy 210 AMSYAQGHFSNAPGTATAEAAAYSLAVGLAAQATKQSSIAV---GSNAKANAF----- 259
Db 196 IAVNGGFFINASKRATLTTCQPOQYQ-AGDLGFKIRGNVVIAGHGLDARDTDFRILSYH 254
Qy 260 -----ATAIGNTVNVNLGKGVAGFGSQILDRDNT-DASAYVPL-----GK- 301
Db 255 SKIDAPWQDVRVAVQNDVAVATGNAHSPILNNAANTSNNTANNNGTHIPLEAIDTGL 314
Qy 302 -----TLADQYKATROGDSDFISIGNSNNNNSIRKLIIVG-----AGSRDTD 346
Db 315 GGMVANKITLITAEQAGIRNQG---OLFA--SSGVAIDANGLRVNSGTPMAANAKDTD 369
Qy 347 AVNVAQLKLVEELANKITFKGDGDNNSNSVERGLNLTIKGAQTNALTEANIGVVTD 406
Db 370 -----NTAEHKVNIRSQGVENS----- 387
Qy 407 GNLKVKLAKELTGLTFSVATNKITVSTNTNNNAELQSGLTFTSPITGKTDTKTVSIDD 466
Db 388 -----TAVSQOQGTQIHQSQIQNTGTLSSGEILLHNSGSLKNETSGTIEA 432
Qy 467 LKFTDNSNIAATKG---TTRITKKKIGFACTND-----GVDESKPYLDNKLKVGSTLN 518
```


Db 433 ARLAIDTDLNNOGKLSQTSOKLHIDAQCKMDNRGRMLQDTAPTASNGS---SNOTCN 489
Qy 519 SGLTVNNTGNKQIOVGANGKFAFVANNVANTSATVCARITEEKIGFAGTNGVDQ 578
Db 490 SYNASFSTTTTPTATG-TGTAIVSINISNITPTFAD-GTIR-----THGALDNS 537
Qy 579 APYLDKERLAVGVEIITDSGI-NAGNHKITGLT-----NGTANTDAVTKOLKDA 628
Db 538 GSIAN-----GQTVSAQOGLNAGQIDHOLNAGSAFNDHNGTIIISDAVHIQ----- 587
Qy 629 KPTLNAGDGISINNSNODLVDSGNITPTIYNSVTKL----- 668
Db 588 -----AG-SLNQN-----GNITT-RQLEIETDQLDNNAHGLLSAETADLAIVS 631
Qy 669 --NSNGTSGNNKFSVNAHD-----NNSLVTAKDLA-----DYLKNVNETADSAL 711
Db 632 LNNNGEATNQOLI--IHGQOQSTAVIDNTNGTIQSGRVDVAIOAKSLSNNGTAAADNKL 689
Qy 712 -----PSFKVQNGDNNATVVGKDTNGKTFNTLKLKGENGVNITNRACTVTFGIDQS 766
Db 690 DIALQDDFVVERNIVAGNELSL--STRGSLKNSHTLOAGKRIRKAKNLDNAAGNI-QS 746
Qy 767 NGLTTPKLTIVGSDTN-GNRLVIEQVPSADGNSTK-----NIIKGLSPTLPSIASPSGR 818
Db 747 GGT-----DIGTQHMLNWRGLI-----DGQOTKIQAGOMNIGTG-----RIYGD 787
Qy 819 NIALGNT-IEEKDKSN-----AASIDVLNAGNKNKDKDFVSTYDVFIDGNATTAT 874
Db 788 NTAIAATRLDNQDENGTAATAARENLNLGIGQLNRENSLIYS-----GNDMAVG 838
Qy 875 VTYDQANQTSKAVDYVNDKFIETLGDNGKKGKOLGVKFI-----KLTETSTNGNAT 925
Db 839 GALTNGQATGAQRIHNAAGTIEAA--GKMLRGEKVLHNTNEHLKQLVETGRE-HIV 894
Qy 926 TSTDDHIALVKAS---DIAGNL-NLAEIHTTKGTANTALQFTVVKYVDENDKADDT- 980
Db 895 DYEAFGRHELLREGTQHELGSWVSDESHLTPDCAAHENWHKYDYBKVTQKTQVTA 954
Qy 981 -----NAITV-KRD--GTSKQVNTLKLKGNGLDIDTKDG---TTFVG---INTQSG 1024
Db 955 PAKIISGNDLTIDGKRVFNTDSOI-----IAGGN-LIVQTEKDLGNEQTFGEKKVSENG 1009
Qy 1025 L-----KAGDST-----TL-----N 1034
Db 1010 KLHSYREXHKGRDSTGHSEQNVTLPETIIRNISLSGFAYESHKALSHHAPSQOTELPQ 1069
Qy 1035 NGLSINTKNTASNEQIOVGADGVKAFMNVNNGVAGIDGTTRIT--RDEIGFTGTNGSLDK 1092
Db 1070 SNGISLPTYSNSTPPLPSSSLYIINPVNKGYL---VETDPFANYROWLGSYMLDSLKL 1126
Qy 1093 SKPHLSK---DGINAGGKITTNIQSGEIAKNSHDVATGKKIYDLKTELENKISSAKTAQ 1149
Db 1127 DPNNLHRLRGDGYE--QRLNEQIAEL--TGHRRLDG---YONDEEQFKALMDNGATAA 1179
Qy 1150 NSLHEFSVADEQGNFNTPNPSYDTSK--TSDVITTFAGENGITTKVNGVVRVGDIDFK 1208
Db 1180 RSM-----NLSVGIALSAEQVAQTSIDVILVQKE-----VKLPDGGTQ 1218
Qy 1209 GLTTPKLTIV---GNNGKGIIVINSONQNTITG-----LSNTLA 1244
Db 1219 TVLVPQVYVRVKNGDIDGKCALLSGSTQINVSGLSKNSGTIAGRNALIINTDTLONIGG 1278
Qy 1245 NVTNDKGSVRTEQ-----GNIIKEDKTRAASIVDVLNAGFNQGNCEAVFVSTYDTVN 1300
Db 1279 RIHAQSAVATODINNIGGLMSAEQTL-----LUNAGNNI----- 1314
Qy 1301 FANGNTTAKVTYDDTSTKSVWYDVNVDDTTIEVKDKIKLGKVTTLTSTGTGANKFALS 1360
Db 1315 --NSOSTTASSONTQGSST-----YLRMAGIYI-----TGKEKGLVA 1350
Qy 1361 NOATGDALVKASDIVAHNLNTLSGDIQAKGASQANNNG-----YVDAD 1404
Db 1351 AQAGKD-----INIAGQISNOSQOTRLQGRDINLDVTQTSKHQATHEDAD 1399

Qy 1405 GNKVIYDSTD--NKYYQAKNDGTVDKTEVAKDKLVAQAQTPDGTILA-----OMNYSVI 1457
Db 1400 -NHVIRGSTNEVGSIOQTKGDTILLSGNNL--NAKAAEVSSANGTLAVSAKNDINISAGI 1456
Qy 1458 NKEQVNDANKQGINEDNAFVKLEKAASNKYKNAAVTVG-----DLNVAQTP 1508
Db 1457 NTHVDDASKHTGRSGGNKLVITDKAQSHHETAQSSTFEKGQVVOAGNDANILGSNVI 1516
Qy 1509 TFAGD-----TGTAKK-----LGETLTIKGGQTDNKL 1537
Db 1517 SDNCTQIOAGNHVRIGTITQOSQSEYHQTKSGLMSAGIGFTIGTSKNTQENOSQSEH 1576
Qy 1538 TDNNIGVAVAGTGTVKLAKDLTLNLSVNA---GGTKIDEXISFVDANGAKANTP--- 1591
Db 1577 TCGTVSLKGDIT--TIVAGKHVEIGSTVSSPEGNNTIYAQSIDIAAHNKLNSNTQTY 1634
Qy 1592 -----VLSANGLDGLKVIS-----NVGKGTKD-----TDAANV--QQLNEVRNLLGL 1632
Db 1635 EOKGLTVAFSPVTDLAQAATAVAOSSKQVGSKNDRVNAMAAANAGHQAQVOTCKSAQNL 1694
Qy 1633 GNDNADGNQVNIADIKDPNSGSSNR--TVIKAGTV-----LGGKNNDEKLTATGGV 1684
Db 1695 ANGTNNAKQVSI-----ITYGEQONROTTOVQAQAQASQIQAGK-----TTLIATGAA 1745
Qy 1685 Q-----VGVDKDGNA-----NGDLSNVVKTOKDGSKKA-----LLATYNA 1720
Db 1746 EQSNINIAGSDVAGKAGTILADIADNITLQSAEQSN---TERGQNKSGAGNAGAVSEGO 1801
Qy 1721 AGQTNVVT--NPAEALDRINEQIRFFHVNDGNOBPVQVGRNGIDSSASGKHSAVIGFO 1778
Db 1802 GWSLGVTAGNVGKGYGNGDSITHRSHSHIGDKGSQTLIQ--SGGDTTIKGAQVRGKQV 1859
Qy 1779 AKA-----DGEAAVAIGROTQAG---NOSIAGDNAQATGDSIAI 1816
Db 1860 VNANKLSIOSVDRETYQSKQNASAQVTVGYGFSAGGYSQSKIRADHVSVTEQSGIYA 1919
Qy 1817 GTG--NVVAKGHSATGD--PSTVKADNSYVGNNNQFTDATQT--DVPFGVGNNTVTVES 1870
Db 1920 GEDGYQKVGNNHDLKGGIITSTQSAEDK---GKNRFOTATLTHSDI---KNHSQYKG 1971
Qy 1871 NSVALGNSAISAGT--HAGTQAKKSDGTAGTTTACATGTCTVKGFAGOTAVGAVSVCSAG- 1928
Db 1972 ESFGLGASASISGRTLGQGAQNKPNKHILTSAADKNSASSVGT--GSDSDSQSSITKSGI 2030
Qy 1929 -----AERRIQNVAAAGEVSATSDAVNGSOLYKATQSIANNATN-----ELD-- 1969
Db 2031 NTRNIQITDEAAQIRLTGKTAQTKADIDTNTVTDPAERHSGSLKNTFNKEAVQSELDLQ 2090
Qy 1970 -----HRIHQENKANAGISSAM-----AMASMPQAVIPGRSMVTCGIATHNGGAV 2016
Db 2091 RTVSQDFSKNVQQAQNTIENQHLDKLADKADKAAETAAEAALANG--DMETAKRKAHQDAA 2149
Qy 2017 AVGLSKLSDN--GQWVFKINGS---ADTQCHVGAAYGA-----GPHF 2053
Db 2150 A-----KADNNQOQKVLINMLASGLAAPTQSGAGIAAATASPASVAYIGQHF 2196

RESULT 13

Q92HL0 PRELIMINARY; PRT; 4919 AA.
ID Q92HL0:
AC Q92HL0:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE LARGE SUPERNATANT PROTEIN 2.
GN LSPA2.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_Taxid=730;
RN [1]
RP SEQUENCE FROM N.A.

RC	STRAIN=35000;
RX	MEDLINE=99030326; PubMed=9811662;
RA	Ward C.K., Lumbley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT	"Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.";
RL	J. Bacteriol. 180:6013-6022(1998).
RQ	EMBL: AF057696; AAC79761.1; --
SQ	SEQUENCE 4919 AA: 542602 MW: 5779201455CA69A0 CRC64;

Query Match

Best Local Similarity 19.9%; Score 540.5; DB 2; Length 4919;

Hatches 503; Conservative 331; Mismatches 913; Indels 785; Gaps 122;

QY	2	NHIKYVFNKATGTFMAVAEYAKSHSTGGGS-----CATQGVGSVRT	43
DB	3	NKRYKLIFSVMKNCVPAENIKSASNGSGSSKIAEDQEDEEPDLSLSPSSIH	62
QY	44	LSPARIAALAVLVIGATLNGSAVAOQTITTKIEGTQNKINNTLKGDALATGEASIPGL	103
DB	63	LGLHNHSPKFVK--GRNLSVVLLSLPAMQVWADSSNAIVDH-----SHGARQTAVER	115
QY	104	SKAQGSO-AIAGSVKPD-----PNNGSGNVGHAKGNESIAIGDDVLAEGBASIAIGS	157
DB	116	DPNKGKEKVVIINTAKPEDGISDNHFSEKFNPSAVFNSII-----REGNSQL-VG-	166
QY	158	DDLPLPNLDLNKEFHKLKHIGHEILKKIQOTSTDGKIYRRTRAQGHASTAVGAMSQAQH	217
DB	167	--LLGENKNLGSOAAKTIFNQ-----	185
QY	218	FSNAPGYTAAEAYSLUAVLAQAQATKQSSIAVCNSKANAFATAIGGTVNVNLGRVA	277
DB	186	-----VTGDQESKISGGLEVFGERADLFIPNP-----GVTNLGVKTIINTDRFA	230
QY	278	-----LGFSGOILDORDN--NTDASYVPLGKTLADQYKATRGDSTD----	317
DB	231	STSEVEPHIKQLNVQRGVLIIGDKDVATNCLSHFDVVAKNIQQGKVSIEGDSKPAKLA	290
QY	318	--IFSIGNSN-----NNNSSIRRKII-----NYGAGSRDTDAVNVAQLKVEELAN	361
DB	291	NVIFAAGNLVDYVTRDVRNTPNPKPIIDNTRKONIAISGESNGSMYGRNIFI-----	345
QY	362	RKITFPKGGDNNSNV--ERGLGNLTIKGAQTNALTEANGVYVTDGNGLKVKLAK---	416
DB	346	--VTDKGAGVHQGVFAIEDDI-NILTDDGNSRLNKVYADVVRV-----GKDIELANNQ	398
QY	417	-----ELTGTSVATNKIIVS--NTNNNAELOSGLTSPITGTK---456	
DB	399	IHADQOILNATGHVKLNDCSSVISNNNLGISALNLTLENATVSANNLSFRVNTDTRLNN	458
QY	457	-----TDKTVYSIDGLKFT--NDNSNIATGKTTRIPIKKIGFAGTNDGV	498
DB	459	LSKVSARAADLOSGNLNDKA--SVLAHLKTLNISDNVSLNQSGLSANNLIKIKVPD---	514
QY	499	DESKPYLDNEKLKVGNSLTN-----SGSLTV--NNTTGNKOIOVGANGIK	541
DB	515	----LNLNNSELSANNLTNLSNNITLKNKSFKTAGNWTLMVNTVNTLNNDSELAAANLT	570
QY	542	FATVANNAVNTSATVGTARITEEKIGFAGINDGVEDQAPYLDRERL---KVGREIITDS	598
DB	571	L-NVTWKVNLNDA-----SKUSANKLDLVNDVNTFLNSKSTLSAGELTFKKVKVNTLND	625
QY	599	GINAGNHKITGLTNGIANTDAVTIKOLKDAKP---TEN-----AGDGISINS-----	646
DB	626	ELAANNLSLASHNVTLNNSKLSAQAKADIKAVNLTLNDTELTAKNLTDINSTIITNNGT	685
QY	647	LVDSSGNITTPTYNISVYKTTKLSNGTSGNKNFSVNAH--DNNSLVTAOKDLADLYLNKVE	705
DB	686	IAGIFANITTEKLNKEKALIL-----AEQNLTFTVNGSHVENKGDIVSKOKATV--TF	739
QY	706	TADSALPSFKVONGDN-----SNNAITVGKDTNGKTFNTLKLKGENVNITNRATGTV	759
DB	740	NSDPTSGNSLVNAQNOQLKVNNNFTISOQDD-----ITLIG-----NVTLN-ASGTF	786

QY	760	TFGIDOSNGLTTPK-LTVGSDTN-----GNRLVIEQVPSADGNSTKNIKGLSPTLP	814
DB	787	T-----NSGNLTVTVKTLDVDGIQNFKNGLTVGEDLHKSKTKITN-----	828
QY	815	PSGRNALCNGTIEBKDKSNAASIDDVNLACFNL-----KNN--CKDKDFVSTYDTV	863
DB	829	-DGKLISIKNL-----NISSEADFNTNGTLTGIEALKIATKGNFTNFKKAILANSLL	880
QY	864	DF--IDGNATTATVYDEA---NOTSKVAYDVNVDEKTIETLTDGNGKKQLGVKTIKLT-E	917
DB	881	DISVAEGKCTFNNGTIESGKNLNTWTCAP-LWVDNATI-----RSFGVLNITSGN	931
QY	918	TSTNG-----NATFTS--TDDHALVKASDIAGNLNLTABEIHTTKGTANTALQTF	967
DB	932	VSNNGTLISNERLNTISAANFTNESNGTMASN---GLLNIAIQOGNITKNLTIASRQOLN	988
QY	968	VKKVDENDKADD-TNAITV-----GKD-----GTSKV-----NWLKL	999
DB	989	LTAVADNITNDSNISNKKIAVLHSLGNSLNSKDQVYNLGEIYAGNNTISVKAHQ	1048
QY	1000	KGNGLDIKT-DKDGTVT-----EGINTOSGLKAGDSTTLNNGLSLTKNTASNE	1047
DB	1049	MG-----DITTKTEGGQASYKLYQASNGCHFGNDGSSSYSGD-----LNING-	1091
QY	1048	QIQVGADGVKFMVNVGVGIDGTTTRITRDEIG--FTGTNGSLDKSKPHLSKDGINAG	1105
DB	1092	-----KPADLN-----KLTQVIGRIYAGRLDTFNKS-----NAG	1122
QY	1106	GK-----KIT-----NIOGETAKNSHDVGGKLYIDLKTELENKISS	1143
DB	1123	GKSEIINRGITNVKNKLSDSDVSFFENNMQSKV-----DLYT-KIFEAKSDIE---	1171
QY	1144	TAKTAQNSLHEFSVADQGNNFTVSNPYSSDYDTSKTSDVITPAGE---NGITTKVNGVV	1200
DB	1172	TFKT--NGTHPVYL-----NFKSNNEKKYRNSENTKNFKSIGDLINEALSAPAEIE	1223
QY	1201	RVGIDTOTKGLTTP---KLTGVNNNGKGINVSONQONTITGLSNTLAN-----	1245
DB	1224	AITYSGSSSNVINPVSILAALGNAN-----NSSPHYLTALKHILNGWGQDDLKKQENI	1277
QY	1246	-----VTNDKGSVRTTEQGNIIKDEDKTRAASIVDWLSAGFNLOGNEADVYSTY	1296
DB	1278	KVLKQKWEDEKDKGASKMLD---LYPNTOKERAKIPAGIIRNGNDTISDVESEDFKKY	1334
QY	1297	DTVNFANG---NITTAKVTYDDTSKTSVVYDVNVDDTTEVDKDKLGVKTTILTSTGTG	1353
DB	1335	S--KFGGEWAKNDDTGSYDSYTSKASEKYKKVENVDHKE-NIDEHKLNIKHEITVPGVS	1391
QY	1354	ANKFALSNOATGDALVKASDIVAHL--NTLSGDIQTAKGASQ-----	1393
DB	1392	FEN--LNKNMHDHPDKLGEIDKSIISELLAQPVYTEKSARSODPRVNQDREALDONLY	1449
QY	1394	-----ANNSAG-----YVDADGNKV-----	1408
DB	1450	RTRLSYINQNQLKACYFFNOLODTEDDKLAGIKRIGDNYFEHQHLITRLEKVDADNHLTK	1509
QY	1409	--IYDST-----DNKYQAQN-----DGTVDKTEKAVDKLVAQAQT-----PD	1445
DB	1510	HGLHDIALVKKLIDSASIQAADLNLKVGEALTKEQKDNLBEDIYVYKTEVNGOEVLVPQ	1569
QY	1446	GTLAOMNVKSVINKEQVNDANKKGINE--DN-----AFVKGLEKAASDNKNTKNA	1494
DB	1570	VYLAQKQIEVEKERQGVGTQIRAGIIDVKVDVDRNTGTITAGYAVGLE---AKNKLK--	1624
QY	1495	VTVGDL-----NAVAQTPLTFAAGDTGT-----	1517
DB	1625	--TGD	

QY 141 IGGDVLAEGDASIAIGSDLLYLPKNLDLKNFHLIHHHEILKIKIOTSTGCKIKYRTRA 200
DB 67 AG---VANG-----TPA 75
QY 201 OGHASTAVGAMSYAGHFSNAGFY-ATAEAYS- --AVGLAAOATKOSSTAVGSNAKAN 257
DB 76 RGPQNN--WAFTYG-----GDTVTADVADHIITAINVA-----DTPIGLNTAQN 119
QY 258 AFAATAI--GGN---TVVNLGRGVALGFGSILDRDNTDASAY---VPLGTLADQYKAT 310
DB 120 TVVGSILITCGNLLPVTIINAGSLTL-----NGNNAADAANHGFDAD-----ADNY--- 163
QY 311 RQGDSTDFISIGNSNNNSSIRRKIINVAGSRDRTDANVVAQLKLVLELANRKTIFKGDG 370
DB 164 -----TCLGNIALGGANAA--LIIQSAAPAKITLAG-- 192
QY 371 DNNSNVERGLNLTIKGDA-----OTNALTEANIGV-VTDGNGLVKLAKELTGLT 422
DB 193 --NING-----GGIITVKTDAAINGTIGNATVNVVAGIATLEGAIK-----AT 238
QY 423 SVSATNKITVSNFNNNAELQSGGLTFSPITGKTOKTVVYSDGLKFTNDSNSIATKGT 482
DB 239 TTKLTNAASVLTNVAVL-----TG-----AIDNTT 266
QY 483 RITKKIGFAGTNDGVDESKPYLDNEKLVGNSTLNSGLT-----VNNTTGNKQIQVGA 537
DB 267 -----GVD-----NVGVNLNLN-GALSQVTGNIGNATLALATISVCA 300
QY 538 NGIKFATVANNVANTSATVGTARITEKIGFAGTNDGVDEQAPYLDKERLKVGRVEITTD 597
DB 301 GK---ATLGAIVIKAT---TTKLTDNASAVTFTNPV-----VTGAID 337
QY 598 SGINAGNHKLT-----GLTNGIANTDAVTIKOLKDAKPTLNAGDGI-----SINNSNGD 646
DB 338 NTGNANNGIIVTFTGDSVTGNIGNTALA-----TVNVGAGLLQVGGVVKANTIN 388
QY 647 LVDSSGNITPTYINISVYKTTKLSNGTSGNKFVSNAHNNLSLTKADLADLYLNKVNET 706
DB 389 LTDNASAVTFT--NPVVVVTGAIDNTGNA-----NNGIVTF-----T 422
QY 707 ADSALPSPKVVONGDONSNAITV---CKDNTGKTFNTLKLKGENGVNITNRRAT---GTV 759
DB 423 GDSVTG---NIGNTALATISVAGKATLGGAI-----IKATTTKLTDNASAV 468
QY 760 TFGIDQSNGLTTPKLTVGS-DTNGNRLVIEQVPSADGNSTKNIILKGLSPILPSIASPSGR 818
DB 469 TF-----INPVVVTGAIDNTGN--ANNIGVTFGTGDSVTGNTIGNATLALATISVAGK 518
QY 819 NIALGNTIEEKDKSNAASIDDDVNLKNGKDKDFVSTYDVFDFID--GNATTATVT 876
DB 519 -ATLGGAIKATTT-----KLTDNASAVTFTNPVVVTGAIDTGNANNGIIVT 564
QY 877 YDEANQTSKVAYDVNVVDEKTIETLTDGCKKQKLGKVTIKLTETSTNGNATFTSTDDHALV 936
DB 565 F---TGDSVTGNIGNTALATISVAGKATLGGAIKATTTKLTDNASAVTFTNPVVVT 621
QY 937 KASDIAGNLTALABELHTTKGTANTALQTFVVKVVDENDKADDTNA-----LTVGK----- 988
DB 622 GAIDNTGN-----ANNIGVTFGTGDSVTGNTIGNTALATVNVVAGIATLE 666
QY 989 GTSKGVTLLKKGKGLDIKTDGTVTFGINTQSGKAGDSTTLNNGL-----STKN 1042
DB 667 GATIKATTTKLTNAASVLTNVAVLGAIDNTTGV--DNVGVNLNLNGALSQVTFGNIGN 724
QY 1043 TASNEQIQVAGDGVKPFAMVNVNGVWVAGIDGTTTRITRDEIGFTGTN-----GSLDKSKPHL 1097
DB 725 TNALATISVAGK---ATLGGAVIKA---TTTTLTNDASAVTFTNPVVVTGAIDNTGN--- 776
QY 1098 SKDCIN--AGKKIT---NIOS-GEIAKNSHDVATGGKIYDLKTELENKI---SSTAKTA 1148
DB 777 ANGIATFTGDSVTGDTIGNTALATISVAGKATLGGAIKATTT---TKLTDNASAVTF 833
QY 1149 QNSLHEFSVADEQGNFTVSNPYSSYDTSKTSVITTFAGENGITTKVKN-----GVVRVGI 1204

QY 1525 ---LTIKG---GQDNTNKLTDNNIGVAGTDGFTVKLAKDLTNL-----SVN 1566
DB 1743 DHLHLSLEGEDVNOQTSNULKNARTTGVVKG--DFNTKAGKDLFHRQIDTVTSGTVVSASAS 1800
QY 1567 AGGAKIDEKIGISFVDANGQAKANPTVLSANGLDGGKVISVNGKTKDIDAANVQQLNEV 1626
DB 1801 GGG---QSAGISLTDQGVETVTKATAGANAD-----VTNFMKRTRETSETSLTHRNSEF 1852
QY 1627 RNLLG-----LGNDNADGNQVN-IADIKDPNMSGSSNRTVIKAGTVLGGKGNNDTEKLAT 1681
DB 1853 NALSGLYVMGKADIGGVDDIINDVEVKTPEEIAAEQAAEEAKKA-EVRENEASEETAAK 1911
QY 1682 GGVQGVQDKGNANGDLNPNVVKTKQDGSKKALLATYNAAGQTVNVTNNPAEAI DR----- 1737
DB 1912 ETEE--AENDNVAEKDKTKPKFKLTD---BEIAAAAFETKGEDEFFAAYKAREEEDRKKGF 1966
QY 1738 -INEGIRFFHVNDGNQEPVQVGRNGRIDSSASGRKSHVAIGFOAKA-----DGE 1784
DB 1967 TLSAEQIESTKARDEKETYIELKVGVAEAEAHSAADAISNKARQIIDFQNGLKQDGT 2026
QY 1785 AAVAGTQAGNOQSIATIGDQAATCGDSIATGTCGNVNVAGKHGSAIGDPPSTVKADNSYSV 1844
DB 2027 VAL-----QASDVLNLTGDLAGASAKLKFELST---IEKSRGA-----SDGRSIL 2071
QY 1845 GNNNOFTDQTDVFGVGNITVTESNVALGNSAISAGTHAGTQAKKSDGTAGTTT 1904
DB 2072 GGRNLNLA-ARGGDI--TLNNVETTENSHLSLKARDNVNNGV-TEQKDESNQSLKVTA 2127
QY 1905 GATGTVKGFAGOTAVGASVGSAGERRIQNVAAAGEVSATSDAVNGSOL-YKATQ--SI 1961
DB 2128 GASGGGVWAGGSAG-VSNVSGS---YNESNTTESTSHTNSLLRGSLSRVEAGKDFNL 2182
QY 1962 ANATNELDHRHONENKANAGISSAMAMSPQAYIPGRSMVTGGIATHNG--QGAVAV 2018
DB 2183 ISSNVVDVH-LHL-DVKGDTNVVSKODSYSRBERGVNYSVSGVSTAGGARPNGSVGL 2240
QY 2019 GLSKLSDNGOWV 2030
DB 2241 GVSANENSIV 2252

RESULT 14
Q52657 PRELIMINARY; PRT; 2021 AA.
ID Q52657;
AC Q52657;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE 190-KDA ANTIGEN (ROMPA).
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALISH 7;
RX MEDLINE=941171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kda antigen-encoding gene of Rickettsia
conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
DR EMBL; U01028; AAA17405.1; --
SQ SEQUENCE 2021 AA; 203366 MW; BE943FBA3BDD5C43 CRC64;

Query Match 5.1%; Score 528.5; DB 2; Length 2021;
Best Local Similarity 22.4%; Pred. No. 3e-14;
Matches 465; Conservative 213; Mismatches 684; Indels 717; Gaps 113;
QY 82 INNTLKGDALATGEASIAFGSLSKAGSQATAIGSVKPDNNGS-NGNVGSHAKGNESIA 140
DB 13 IOGGLKAALFTTSTAIMLSS-----SGALGIASVGIATNNNAFSDNVGN--WNEITA 66

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 23:04:48 ; Search time 33595.7 Seconds
(without alignments)
3196.157 Million cell updates/sec

Title: US-09-361-619-10

Perfect score: 6942

Sequence: 1 atgaatcacatctataaagt.....ttggtgcaggtttcaattt 6942

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_bal:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vi:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_vil2:*
- 60: gb_htg1:*
- 61: gb_htg2:*
- 62: gb_htg3:*
- 63: gb_htg4:*
- 64: gb_htg5:*
- 65: gb_htg6:*
- 66: gb_htg7:*
- 67: gb_htg8:*
- 68: gb_htg9:*
- 69: gb_htg10:*
- 70: gb_htg11:*
- 71: gb_htg12:*
- 72: gb_htg13:*
- 73: gb_htg14:*
- 74: gb_htg15:*
- 75: gb_htg16:*
- 76: gb_htg17:*
- 77: gb_htg18:*
- 78: gb_htg19:*
- 79: gb_htg20:*
- 80: gb_htg21:*
- 81: gb_htg22:*
- 82: gb_htg23:*
- 83: gb_htg24:*
- 84: gb_htg25:*
- 85: gb_pr1:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_rol:*
- 95: gb_rod:*
- 96: gb_in4:*
- 97: gb_pr10:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6942	100.0	6942	10	AX079922 Sequence
2	2939.2	42.3	6159	10	AX079920 Sequence
3	2852.4	41.1	62909	10	AX067457 Sequence
4	2590	37.3	6141	10	AX079918 Sequence
5	2590	37.3	6972	10	AX079917 Sequence
6	2580.6	37.2	6973	9	AR040716 Sequence
7	2580.6	37.2	6973	10	AX079913 Sequence
8	2485.8	35.8	5976	10	AX079914 Sequence

[illegible]

Qy	1141	caagcggacaagctaacccgataataataaacattgggtggtgtaacagagataataataactgggt	1200
Db	1141	CAAGCGACAAGCTAACCCGATTAATAATACATTTGGTGTGGTAACAGAGATAATAATACTGGT	1200
Qy	1201	ctgaagattaaacttctgtaaaaacctaagcgtgtcttgaacagttagcaccaaaaaacctta	1260
Db	1201	CTGAAGATTAACTTCTGTAAAAACCTAAGCGGCTCTTGAACAGATTAGCACCAAAAACCTA	1260
Qy	1261	accgcagcgagaaagctacggttaggttagtggtaataaacacgcgtgagctacaaagcgtt	1320
Db	1261	ACCGCCAGCGAGAAAGTTACGCTAGGTAGTGGTAAATAACACCGCTGAGCTACAAAGCGGT	1320
Qy	1321	ggtttaacctttaccccaacaaataacaagcacagacagaaaaacggtctatggcaactgat	1380
Db	1321	GGTTTAACCTTTACCCCAACAAATGCAGACAGACAAACCGTCTATTGGCACGTGAT	1380
Qy	1381	gggcttaagtttactgataattctaactcagcgcaacttgaagatactactcgtataccaaaa	1440
Db	1381	GGGCTTAAGTTTACTGATAATTCTTAATACGCGACATTGAAGATACTACTCGTATCACCCAAA	1440
Qy	1441	gataaaattggtttttagcaataaagcgtggtagcttgatgtaaaacaaacotttatcttgat	1500
Db	1441	GATAAAATTGGTTTTAGCAATTAAGAGCTGGTAGCTAGTTGATGAAAAACAACCTTATCTTTGAT	1500
Qy	1501	aaagacaagctaaagttggcaacgacgaccctaaacaacggtgcttgactgactgtaataac	1560
Db	1501	AAGACAAAGCTAAAGTTGGCAACGACGCCCTTAACAACGGTGGCTGACTGTTTAATAAAC	1560
Qy	1561	accattggtgttagcaataaaaceaatcccaagtcggtgctgattgcaattaaatttgcgat	1620
Db	1561	ACCAATTGGTGTAGCAATAAACAAATCCAAGTCGCTGCTGATGGCATTAAATTTGCCGAT	1620
Qy	1621	gtgaattgttaattgatacgaatgcgcgaataatctggcactactcgtatacogaagaggaa	1680
Db	1621	GTGAATCTTAATGTATCAAAATGCCGCAAAATTCGGCACTACTCTGTTATTCGGAAGAGGAA	1680
Qy	1681	attgcttctgctgactgtagtgtaagcttgataaaaaagtcaccatatttggataaaaaa	1740
Db	1681	ATTGGCTTTGCTGATGCTGATGGTAAGTTGATAAANAAGTCACCATNTTGGATTAANAAA	1740
Qy	1741	caactcaagtcgggtgggtgttaaaattaccaagacagtggcatttaactcaggtgatcaa	1800
Db	1741	CAACTTCAAGTGGTGGTGTAAAAATTACCAAGACACTGGCATTAAATGCAGGTGATCAA	1800
Qy	1801	aagatcagtaattgtaaagatgcaacgacgacgataccgtagcagtcacttataaaacagctt	1860
Db	1801	AAGATCAGTAATGTTAAAGATGCAACGGACGATACCGATGAGTCACCTTATAAACAGCTT	1860
Qy	1861	aaacaagtccaacaagacgcgacggtgccttacaagcttctattcgtgatgaaaaa	1920
Db	1861	AAMCAAGTCCACAGACGCCGACGGTGCCCTTACAAAGCTTCTCTATTCTGGTATGAAAAA	1920
Qy	1921	ggtcaggaatttcagattagtaacttgatttctaattggtataccccaaataacctttgag	1980
Db	1921	GGTCAGGAATTTACGATTAGTAACTTGATTCTTAATGCTAATACCCCAAAATACCTTTTGAG	1980
Qy	1981	accalcacctttgcaggtgaaacggcatalcagtalccagatagacatagccaaggtataa	2040
Db	1981	ACCATCACCTTTGCAGGTGAAACCGGCATCAGTATTCAGCAATGACATAGCCAAAGGTAAA	2040
Qy	2041	gtcaaatgtgatttagaccaatcaatgggtctcaccacgacctgaagctgacgtgggtagc	2100
Db	2041	GTCNAAGTTGGTAATTGACCAATCAATGTGCTTCACACGCGCTTAGCTGACCGTGGGTAGC	2100
Qy	2101	gataaagatggttaaaactcaattgggtatttagcgaagtgggtcaggttaacgacaccaaa	2160
Db	2101	GATAAAGATGCTAAACTCAATTTGGTTATTGAGCAAGTGCCTACGGGTACGACACCCAAA	2160
Qy	2161	acaatcattagaggtatgttcccaacaactgcctagcatataccaatgcaggtgcggtacgc	2220
Db	2161	AACATCATATTAGAGGATTGTCCCAACACTGCCATGCGTAGCATTTAGCAATTGACGATGAGGTGGCGTACGC	2220

Qy	2221	acacagaaacagggcaatacaatcaccagcgacgaagacaataccaagcgcagtcac	2280
Qy	2221	acacagaaacagggcaatacaatcaccagcgacgaagacaataccaagcgcagtcac	2280
Qy	2281	ggtgatatataataacagcgtttaacctcaaaataatagcaactcgttgcttgctc	2340
Qy	2281	ggtgatatataataacagcgtttaacctcaaaataatagcaactcgttgcttgctc	2340
Qy	2341	tcacattataacactgtgactttatcgatggaatgcacacacgcgttaaggttaacttac	2400
Qy	2341	tcacattataacactgtgactttatcgatggaatgcacacacgcgttaaggttaacttac	2400
Qy	2401	gatgaacaacatcaaacagagtaagaacttatgtgtcaatgtggtatgagaaaaacatt	2460
Qy	2401	gatgaacaacatcaaacagagtaagaacttatgtgtcaatgtggtatgagaaaaacatt	2460
Qy	2461	gaactcacagcgataatggcagaacaaataatggcgtcaaaacacacacactgacc	2520
Qy	2461	gaactcacagcgataatggcagaacaaataatggcgtcaaaacacacacactgacc	2520
Qy	2521	acaacaaatgctaagtgaagaacaaactttagtaaccacgataacagatgccttggt	2580
Qy	2521	acaacaaatgctaagtgaagaacaaactttagtaaccacgataacagatgccttggt	2580
Qy	2581	aacgccaagaacatcgccgaataatcaaacacccctagccaagggaattcacacacacaa	2640
Qy	2581	aacgccaagaacatcgccgaataatcaaacacccctagccaagggaattcacacacacaa	2640
Qy	2641	ggcagcagcagacacccctacaaactttaagtcacaaagacggtgcacactgatgac	2700
Qy	2641	ggcagcagcagacacccctacaaactttaagtcacaaagacggtgcacactgatgac	2700
Qy	2701	gaacacatcacccgtgggtgaagatggtcacacaaacgcgaagacgcgtcaacactctaaa	2760
Qy	2701	gaacacatcacccgtgggtgaagatggtcacacaaacgcgaagacgcgtcaacactctaaa	2760
Qy	2761	ctcaagagtgaacacggcttaacggttggtacacataaagatggtacactttggc	2820
Qy	2761	ctcaagagtgaacacggcttaacggttggtacacataaagatggtacactttggc	2820
Qy	2821	atlaacacccaagcggctttaaagcggcgacacacacactcaaaagaagatggcttg	2880
Qy	2821	atlaacacccaagcggctttaaagcggcgacacacacactcaaaagaagatggcttg	2880
Qy	2881	tcctataaaaccccctagtagtaacgaacaaatcccaagtcggtgctgaaggcgttg	2940
Qy	2881	tcctataaaaccccctagtagtaacgaacaaatcccaagtcggtgctgaaggcgttg	2940
Qy	2941	gccaaagttgataaagggttaattcaagcactggtatggtgcaagaagcgttatcaccaa	3000
Qy	2941	gccaaagttgataaagggttaattcaagcactggtatggtgcaagaagcgttatcaccaa	3000
Qy	3001	gatcaaatgtgcttactgggctaagtgcactgtatgataccacaaaaccccaactaac	3060
Qy	3001	gatcaaatgtgcttactgggctaagtgcactgtatgataccacaaaaccccaactaac	3060
Qy	3061	aaagacaagcttaaaagggtgaagttgaattaccaacactggcattaaacagcagtggt	3120
Qy	3061	aaagacaagcttaaaagggtgaagttgaattaccaacactggcattaaacagcagtggt	3120
Qy	3121	aaaagattaccacattcaatcaggtgatattaccacaaacagcaatgatgctgtgaca	3180
Qy	3121	aaaagattaccacattcaatcaggtgatattaccacaaacagcaatgatgctgtgaca	3180
Qy	3181	ggcgtcggtgttatgattttaaaacccgaactttgaagaacaaatacaacagtgctaaa	3240
Qy	3181	ggcgtcggtgttatgattttaaaacccgaactttgaagaacaaatacaacagtgctaaa	3240
Qy	3241	acagcacaaactcattacagaattctcagtagcagatgaacaaaggttaatacactttacg	3300
Qy	3241	acagcacaaactcattacagaattctcagtagcagatgaacaaaggttaatacactttacg	3300
Qy	3301	gttagtaaccccttactccagtttatgacacccctcaaaagacccctgtatgctacacctttgca	3360

Db	4381	GGTGGCAAAACAGACACCAATAAGCTAACCGATAATNACATCGGTGGTAGCAGGTACT	4440
Qy	4441	gatggttcaactgcataacttgcataagacctaaccaacttcaacagcgcttaatcgaggt	4500
Db	4441	GATGGGCTTCACCTGTCRAACTTTGCCAAAGACCTTAACCAATCTTAACAGCGTTAATGTCAGGT	4500
Qy	4501	ggcaccagaaattgtagaaaaaggcatctcttttgaagcacaacaggtcacaagcacaagca	4560
Db	4501	GGCACAGAAITTGATGAAAAAGGATCTCTTTTGTAGACGCAACACGGTCAAGGCCAAGCA	4560
Qy	4561	aacacccctgtcttaagtgcacaatggctggacctgggtggcgaacgcatcagtaacatc	4620
Db	4561	AACACCCCTGTGCTTAAGTGCCTAATGGCTGACNCTGGGTGGCAACGCATCACTAATACATC	4620
Qy	4621	ggctgagctgttgatgataacgatcggtgaaactttaagcagtttaagtgaagttgcacaa	4680
Db	4621	GGTGCAGCTGTTGATGATTAACCATCGGTGAACCTTAAGCAGTTTAATGAAGTTGGCCAAA	4680
Qy	4681	acggtcaacaaccctaaacaacaaagtaacacagtgctgctcatcacctcttggtglaacc	4740
Db	4681	ACGGTCAACAAACCTTAACAAACCAAGTAACCTCAGGTGGCGCTCATTTACCTTTGTGGTAACC	4740
Qy	4741	gatgccaattggcgaagcccatcaattggcaccgtagcgaagcccccaaaagccatcaagggc	4800
Db	4741	GATGCCAATGGCAAGCCATCAATGSCACCGATGGCAAGCCCAAAAAGCCATCANAGGGC	4800
Qy	4801	ggcgatggtaataactatcacgcccaacgccaagcggtacacctggtgacaagaatggcaag	4860
Db	4801	CGCGATGGTAAATACTATACAGGCCAACGGCCACGGGTACCTGTGGACAAAGATGGCAAG	4860
Qy	4861	cccatcacccgatcgcgacaacttgcgaacttggcagctcatggaaccccccttgatgca	4920
Db	4861	CCCATCACCGATGGGACAAACTTGGCAATCTGGCAGCTCATGGCAAAACCCCTTTGATGCA	4920
Qy	4921	ggatcaatcaagtgggtgcgaacctaggggcgaactcagatgccatcacctcaaccaaatc	4980
Db	4921	GGTCATCAAGTGGTGGCAAGCCTAGGGGGGCMCTCAGATGCCATCACCCTAACCAACATC	4980
Qy	4981	aagtcacctttgcccacaaattgacacaccaaacacaggttaatgccaatgcaagggcgaagcc	5040
Db	4981	AAATGCCATTTTGGCCAAATTTGACACACCAACACAGGTAAATGCCAATTCAGCGGCAAGCC	5040
Qy	5041	caaatcttgcacagcctatcagcagcagacaagaataatgctgccagtgcaaatgtg	5100
Db	5041	CAAAATCTGCCAGCGCTATCAGCAGCAGACAAGAAATGCTGCCAGTGTCAAGATGTG	5100
Qy	5101	ctaaatgtaggctttaacttgcagaccaaatacaaatcaagtggactttgtcgaagcctat	5160
Db	5101	CTAAATGTAGGCTTTAACTTGCAGACCAATACACAATCAAGTGGACCTTTGTCAAGGCCAT	5160
Qy	5161	gataccgtcaacttctgtcaatgtgtacaggtgcgagacatcacagcgctgcgtagctgat	5220
Db	5161	GATACCGTCAACTTTTGTAAATGGTACAGTGGCCAGCATCAACAGCGTGGTACTGCTGAT	5220
Qy	5221	ggcacgataglaacatcacctgctcaaacaccgcttagcagaccgatgtagtggaat	5280
Db	5221	GGCACGATGAGTAAACATCACCTGTCAACACCGCCCTTAGCAGCGACCGATGATGTGCAAT	5280
Qy	5281	gtgcttatcaaaagccaaagatggtgaattctacaaaagcagacacctcatgccaaagcgc	5340
Db	5281	GTGCTTATCAAGGCCAAAGATGGTAAAGTTCTACAAAGCAGACGACCTCATGTGCCAAGCGC	5340
Qy	5341	tcactaaaagcaggcaaatcaggcagtgatgcaaaaactccaaactggcttaagccttgtt	5400
Db	5341	TCACTAAAGCAGGCAAAATCACGCCAGTGAATGCCAAAACTCCAACTGGTCTTAAGCCTTGTT	5400
Qy	5401	aaccccaatgctgtaaaagcagtcacaggcgaatgcagtggtcttataataacttatcaaa	5460
Db	5401	AACCCCAATGCTGGTAAAGGCAGTACAGCGCATGACAGTGGCTTCATTAAGCCTTGTT	5460
Qy	5461	gcggatatttaaaatccaaagatggtacaactctaccacagtaagctctgtagggcatcagt	5520
Db	5461	CGGGATTTAAATCCAAAGATGATCACTACTACACAGTAAAGCTCTGATGGCATCAGT	5520

[illegible]

Qy	1327	acctttaccaccaacaacaatgaagcagcacagacaaacacgctctatgacctagctgggctt	1386
Db	1342	ACCTTTTACGCCCAATAACACAGGTGACAAAACAGATAAAACCGTCTACACAGCATTTGATGGATTG	1401
Qy	1387	aagtttac---tgataattctaaacggcaccttgaagatactactactcgtaoccaagat	1443
Db	1402	AAGTTTACTAATGATAGTAAATAGTATAGCAACATAAAGGTACTACTCGTATTACCAAAAAG	1461
Qy	1444	aaaattggtttagcaataaagctggtcacagttgatgaaacaaaacaccttatcttgataa	1503
Db	1462	AAAAATTGGTTTTTGGTGGTACTAATATGATCGAGTTCATGAAAGCAAAACCTTATCTTTGACAAC	1521
Qy	1504	gacaagctaaagttgcaacagcacccctaaacacacggttggcttgactgctgtaataacacc	1563
Db	1522	GAAAGCTTAAAGTTTGGCAACGACCCCTTAACACAGTGGTAGCTTGACTGTTAATAACACC	1581
Qy	1564	attgggtgtagcaataaacaatccaagtcggtgctgatggcattaaatttgcgcgattg	1623
Db	1582	ACTGCT-----ATAAAACAATCCAAGTCGGTGCTTAATGSCATTTAATTTGCCACAGTC	1635
Qy	1624	aatgtaattgatacaaatgcc---gcaaaattcggcactactcgatataccgaagagaa	1680
Db	1636	GCATAATAATTTGGCAATATACCTCAGCAACAGTTCGGCACTGCTGCTATTACCGAAGAGAA	1695
Qy	1681	attggcttggctgctgatggtgtaaaattgataaanaagtcaccatatttggataaaaaa	1740
Db	1696	ATTGGTTTTTGGTGGTACTAATGATGCAGTTGATGAACAAGCACCATATTTTGGATTAAGAA	1755
Qy	1741	caactcaagtggtggtgttataaattaccaaaagacagtgccattaaatgcaggtgataa	1800
Db	1756	CGACTTAAAGTGGGTGGTGTGAAATTTACCACAGATAGTGGTNTTATGCTGGTAAATCAC	1815
Qy	1801	aagatcagtaattgaaagatgcaagcagatccgataccgataccgatacttaataaacagtt	1860
Db	1816	AGAATTTACCGGACTTACTAATGGTATAGCAAAATACCGATGCGGTATACCATCAACAGCTC	1875
Qy	1861	aaacaagtccaaacagacgccaggtgcccataaagctctctattcgatgatgaaaaa	1920
Db	1876	AAAGACGCCAAGCCTACTTTAAACGCAGCGCATGGCATCAGTATTAAATAGTAAATTAACGGG	1935
Qy	1921	ggtcaagaatttacgattagtaactgtatgc-----	1952
Db	1936	GATCTTAGTGTAGTAGTAGTGCAATATTACCACCCCACTTATAACATTAGCTGTAAMACC	1995
Qy	1953	-----	1952
Db	1996	ACTAAGCTTAACAGTAAATGGCACCACTGGTAAATAATATAATTTAGTGTTAGTAACTGCTCAT	2055
Qy	1953	-----	1952
Db	2056	GATTAACAATAGCTTAGTTACCGCCAAAGATTTGGCAGACTATCTAAATAAAGTCAATGAA	2115
Qy	1953	-----	1952
Db	2116	ACGGCTGCAGTGTCTTACCAAGCTTTAAAGTCCAAACCGGTGATTAATAGCAACAACGGC	2175
Qy	1953	-----taatgtaatacccaaatcacctttgagaccatcacctttgcaggtgaa	2001
Db	2176	ATCACCGTGGGTAAAGATACAAACGGCGACACCTTTCAACACCTTTAAACATCAAGGTGAA	2235
Qy	2002	aacgcatcagatcacgcaatgacatagcacaaggttaaagtcgaattggtatgtgacca	2061
Db	2236	AACGGTGTATAATTACGACCAATAGAGCCACAGGTACAGTTACCTTTTGTCATTTGACCAC	2295
Qy	2062	atcaatgggtctcaaccgccttaagctgacagttgggtgagcataaagatggtaaaactcaa	2121
Db	2296	AGTAATGGTCTCAGCACGCCCTTAAGCTGACCGTGGGTAGCATACCAATGGTAA---TCGA	2352
Qy	2122	tgtgttattgacaagtggctag-----cgttaacgacacccaaaaacattcattaggga	2175
Db	2353	TTGGTTATTGAGCAAGTCCCTTAGCGCTGCACGGTAAACGACCAAAAAACATCATTTAAAGGA	2412
Qy	2176	ttgtccccaacacatgccttagcattaccaatcaggtggcgtacgcacacacgaacagggc	2235

Qy	4375	atcaaaagtggtggaacacagacacccaataagctaaacccagataataacatcgtgtgtgtagca	4434	Qy	5455	tcaaaagcgtgtatttaaatccaaagatggtacaactactaccacagtaagctctgtatgac	5514
Dd	4579	ATCANAAGTGGGNCNACACACACCAATTAAGCTAACCGATAATAACATCGGTGTGTAGCA	4638	Dd	4807	-----	4806
Qy	4435	ggtaactgattggttgaactgtcaaaacttgccaagacccttaaccaatcttaacagcgttaat	4494	Qy	5515	atcaglatccaaagcaaaagataaacagcagcatcacccctaaagcaaaagatgggctgaatgta	5574
Dd	4639	GGTACCTGATGGCTTCACTCTCAAACTTGCCAAAGACCTTAACCAATCTTTAACAGCGTTAAT	4698	Dd	4807	-----	4806
Qy	4495	gcaggtggcaccagaattgatgaaaaggcatctctctttgtagacgcacaaacgggtcaagcc	4554	Qy	5575	ggcgggtaagtcctacagcaatgTgggtTaaagggcacaaaagacacaccgacgctgccaatgta	5634
Dd	4699	GCAGGTGGCACCACCAAAATGATGAAAAGGCATCTCTTTGTAGACGCAACCGTCAAGCC	4758	Dd	4807	-----	4806
Qy	4555	aaagcaaacacccctgtgtaagtgcgaatgggctggacacttgggtggcaaacgcatacgt	4614	Qy	5635	caacagttaaacgaagtacgcacactgtgtgggtcttgggttaactgctgtaagtataacgct	5694
Dd	4759	AAAGCAAAACACCCCTGTGCTAAGTGCAATGGCTGGACCTGGCTGGC-----	4806	Dd	4861	CAACAGTTAAACGAAGTACGCAACTTGTGGGTCT-----TGGTAATGATATACGCT	4911
Qy	4615	aacatcgggtgcagctgtgatgataacgatgcggtgaaactttaagcagtttaagtt	4674	Qy	5695	gacgcgaatcaggtTaaacattgcgcacatcaaaaaagaccccaaatcaggttcatcatct	5754
Dd	4807	-----	4806	Dd	4912	GACGGCAATCAGGTAAACATTTGCCGACATCAAAAAAGACCCAAATTCAGGTTTCATCATCT	4971
Qy	4675	gccaaaacggtcaacaacctaacaacaaagtaactcaggTgcgtcatcaccctttgtg	4734	Qy	5755	aaccgcactgtcatcaaaagcagcgacacggttacttggcgtTaaaggttaactacagataccgaa	5814
Dd	4807	-----	4806	Dd	4972	AACCGCACTGTCTATCAAAAGCAGGCACGCTACTTGGCGGTAAAGGTAAATACGATACGGAA	5031
Qy	4735	gtaaccgatgccaatggcaagcccatcaatggcacgatggcaagcccaaaaagccatc	4794	Qy	5815	aaacttgccactggtgtgtgtacaagtggcgtggataaaagcggcaacgctaaacggcgt	5874
Dd	4807	-----	4806	Dd	5032	AAACTTGCACCTGGTGGTGTACAAAGTGGCGGTGGATTAAGNCGGCACGCTAAGGGCGAT	5091
Qy	4795	aaggcgcgcatgtaaaatactatcacgcacacgcgaacgcgaactgtacactgtgacaaaagt	4854	Qy	5875	TaaagcaatgtttgggtTcaaaaccccaaaaagatggcagcaaaaaaagccctgctgcacact	5934
Dd	4807	-----	4806	Dd	5092	TTAAGCAATGTTTGGTCAAAACCCAAAGATGCGACGCAAAAGGCGCTTGTCCGCACT	5151
Qy	4855	ggcaagccatcacccgatgcggaacaaacttgccaacttgccaactggcagctcatggcaaacccctt	4914	Qy	5935	TataacgcgcaggtTcagacccaactatttgaccaaacaccccccgcagagccattgacaga	5994
Dd	4807	-----	4806	Dd	5152	TATAACGCGCGAGGTTCAGACCACTATGTGACCAACCAACCCCGCAGAGCCCTTGTACAGA	5211
Qy	4915	gatgcaggtcatcaagtggTggcaagccttagggcgcaactcagatgccatcaacctaaac	4974	Qy	5995	ataaatgaacaaaggtatcgctcttccatgtccaacgtggcaatcaagagccctgtggtga	6054
Dd	4807	-----	4806	Dd	5212	ATATATGATCAACAGGTATCGGCTTCTTCCATGTCAACGATGCAATCAAGAGCCTGTGGTA	5271
Qy	5035	caagccaaaagtctgccagcctatcacgacgacagcaaaagtaatgctgccagTgtcaaa	5094	Qy	6055	caagggcgtTaaacggcatTgaactcaagtccctcagcgaagcactcagtgccgatatagtttc	6114
Dd	4807	-----	4806	Dd	5272	CAAGGCGGTAAAGCGCATTTGACCTCAAGTGCCTCAGGCAAGCACTCAGTGGCGATAGTTTC	5331
Qy	5095	gatgtgtaaatgttaggtcttaacttgacagcccaatcaccaatcaagtggactttgtcaaa	5154	Qy	6115	caagccaaagcagatggTgaagcccgcttgccatagcagacaaaaccccaagcagggcaac	6174
Dd	4807	-----	4806	Dd	5332	CAGGCCAAGCGAGATGGTGAACCGCGCTTGCCATAGGCAGACAAACCCAAAGCAGGCAAC	5391
Qy	5155	gcctatgataacgctcaactttgtcaatggtacaggtgcccagatcaacaagcgtgcgtagt	5214	Qy	6175	caatccatgcctcctcgtgtgataaacgcacaaagccacggcggtatcaatccatgcctcgt	6234
Dd	4807	-----	4806	Dd	5392	CAATCCATCGCATCGTGTATACGCACAAAGCCACGGGGGATCAATCCATCGCCATCGGT	5451
Qy	5215	gctgatggcacgatgagtaacatcacctgacaggtgcccagatcaacaaccccttagcagatgat	5274	Qy	6235	acaggaatgtggtTaaacaggttaagcactctggtccatcggcagcccaagcactgttaag	6294
Dd	4807	-----	4806	Dd	5452	ACAGGCAATGTGGTAGCAGGTAAAGCACTCTGTTGGTGCATCGCGACCCAAAGCACTGTTAAG	5511
Qy	5275	ggcaatgtcttatcaaaagccaaaagatggttaagttctacaaagcagacacacctcatgcca	5334	Qy	6295	gctgataacagtttacagttggtggttaataacacccagtttatcgatgcctcagaccact	6354
Dd	4807	-----	4806	Dd	5512	GCTGATAACAGTTTACAGTGTGGGTAAATAACAACCACTTTACCGATGCCACTCAAAACCGAT	5571
Qy	5335	aacggctcactaaaagcaggcaaatcagccagtgatgccaaactccaactggcttaagc	5394	Qy	6355	gtctttggtgtgggcaataacacacccgtgagccgaagTaaactcgtgtgccttaggttca	6414
Dd	4807	-----	4806	Dd	5572	GTCTTTGGTGTGGGCATATACATCACCGTGACGGAANGTAACTCGGTTGCCTTAGGTTTCA	5631
Qy	5395	cttgTtaaccccaatgctgtgtaaaagcagTacagcgcagTacagcgcagTgctcttaataactta	5454	Qy	6415	aactctgccatcagtcaggcacacacgcagggcacacaaagccaaaaaaactctgacggcaca	6474
Dd	4807	-----	4806	Dd	5632	AACTCTGCCATCAGTCAGGCACACACCGCAGGCACACAAAGCCAAAAAATCTGACGCACA	5691
Qy	5395	cttgTtaaccccaatgctgtgtaaaagcagTacagcgcagTacagcgcagTgctcttaataactta	5454	Qy	6475	gcaggtacaacccacccacagcaggtgcaacaggttacaggtTaaagcctttgctgacaaaaag	6534
Dd	4807	-----	4806	Dd	5692	GCAGGTACAAACCCACACAGCAGGTTGCCACAGGTACGGTTAAAGCCTTCTGCTGGCAACG	5751
Qy	5395	cttgTtaaccccaatgctgtgtaaaagcagTacagcgcagTacagcgcagTgctcttaataactta	5454	Qy	6535	gcggttggtgcggtctccgtgggtgctcctcaggtgctgaaacgcgcgtatccaaaaatgtggca	6594

QY 3486 cagtaagaatggtcaaaaataccatcacaggaactaagaacaactcttagttaaattaccaa 3545
|||||
Db 46053 CAGTAAAGATGGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAAATGTTACCAA 46112
QY 3546 tgaatggtcaggaacacgactaaagcaagggcttgccaatgacacgacaaacccgtgc 3605
|||||
Db 46113 TGATGTGAGGACAGCGCACTAAGCCAAAGGGCTTGCCGATACACCGCAAAACCCGTGC 46172
QY 3606 cgcagcattggtgatgtctaaacagcaggtcttaacttgcaagggcaatgggtgaagcgggt 3665
|||||
Db 46173 TGCCACGATGGTGATGATTAATGACAGGCTTTAACTTGCAAGGCAATGGTACAGCGGT 46232
QY 3666 tgaactttgtctcaacttaagacactgttgactttaacgtatggtgcaatgcccacacccgtctaa 3725
|||||
Db 46233 TGACTTTGTCTCCACTTATGACACTGTCAACTTTGCCATGTCGCAATGCCACACCGCTAA 46292
QY 3726 ggtgacctatgatgacacgaagcaaaacacagtaaaagtggctctatgctgaatggtgataa 3785
|||||
Db 46293 GGTGACCTATGATGACACGAGCAAAACCACTCAAGTGGTCTATGATGTCAATGTGGATGA 46352
QY 3786 taaaaccattgaagtgaacaagtataaaaaacttggcgtcaaaacccacacacactgaccaa 3845
|||||
Db 46353 TACAACCATTTGAAGTTAAA---GATAAAAACTTGCGTAAAAACCCACCATTTGACCGAG 46409
QY 3846 acaagtgctaaatggttaatgca---accaaatttagtgcgcgcatggtgcagatgccccttgt 3902
|||||
Db 46410 TACTGCGCAGAGTGTCTTAATAAATTTGCCCTAAGCAATCAAGCTACTTGGCGATGCGCTGT 46469
QY 3903 taaagcagtgatgcaccacatcaataaccttggctggcgacatccaaacccgcaa 3962
|||||
Db 46470 CAAGGCCAGTGATGCTGCTCATTAACACCTTATCTGGCGACATCCAAACTGCCAA 46529
QY 3963 aggggcaagcgaagcagcgtcagcaagctatgtggatgctgtggtgcaacaaagttcat 4022
|||||
Db 46530 AGGGCAAGCAAGCAACAACTACGACGCTATGTGGATGCTGTGTCGCAATTAAGTTCAT 46589
QY 4023 ctatgacagtacogataagaagtaactatacaagtcaatgacaaaggtgcaagtggaacaaaa 4082
|||||
Db 46590 CTATGACATFACCGATTAACAAGTACTATCAAGCCAAAATGATGGCACAGCTTGTATAAAC 46649
QY 4083 caaagaattgccaagacaaaactggtgcgccaagcgaaccccgatggcacattggc 4142
|||||
Db 46650 CAAAGAAGTTGCCAAGACAAACTGCTGCCCAAGGCCCAACCCAGATGGCACATTTGCC 46709
QY 4143 tcaaatgaaatgcataatcagtcattacaagaagcaagtaaaatgctgccaataaaaaagca 4202
|||||
Db 46710 TCAAAATGAATGTCAAAATCAGTCATTAACAAAGAAACAAGTAATGATGTCCTAATAAAGCA 46769
QY 4203 aggcataatgaagacaaagcgttttatcaaaagggcttgaacgcgcgcaaaagacaccaa 4262
|||||
Db 46770 AGGCATCAATGAAGACAAACGCCCTTTGTTAAAGGACTTGAANAAGCGCTTCTGTATAACA 46829
QY 4263 aacccaaacgcgcagtaactgtggtgatttaaatgcccgttggccaacacccgtgac 4322
|||||
Db 46830 AACCAAAAACCCGCCAGTAAGTGTGGTGNATTAATGCGCTTGCCCAACACCCGCTGAC 46889
QY 4323 ctittgaggggatacaggacacacgcgttaaaaaactggcgagactttgaccatcaaaag 4382
|||||
Db 46890 CTTTGCAGGGGATACAGGCACACGCTTAATAAATCTGGCGAGACTTTTGACCATCAAAAGG 46949
QY 4383 tgggcaaacagacaccaaataagataaccgataataacatcaggtgtggtgagcaggtactga 4442
|||||
Db 46950 TGGGCAAAACAGACACCAATAAAGCTAACCGATAATACATCGGTGTGGTATGACAGGTACTGA 47009
QY 4443 tggcttcaactgtcaaaacttgccaaagacaccttaaccaatcttaacagcgttaatgacaggtg 4502
|||||
Db 47010 TGCGTTCACTGTCAAACTTGCCCAAAACACCTAACCAATCTTTAACAGCGTTAATCAGGTGG 47069
QY 4503 caccagaattgatgaaaagggcatctctttttagacgcgaacacgggtcaagccaaagcaaa 4562
|||||
Db 47070 CACCAAAATTTGATGACAAAGGCGTGTCTTTTGTAGACTCAAGCGGTCAAGCCCAAGCAAA 47129
QY 4563 caccctctgtctaagtgcgaatgggctggacctgggtggaacacgcacatcagtaaacatcgg 4622
|||||

Db 47130 TACCCCTGTGCTAAGTGCCAATGGGTGGACCTGGGTGGC----- 47169
QY 4623 tgcagctgttgatgataaagatgcggtgaactttaagcagtttaagaaagttgccaaaac 4682
Db 47170 ----- 47169
QY 4683 ggtcaaacacctaataacaaacaaagtaactcaggtgogtcatcaccctttgtgtaaccga 4742
Db 47170 ----- 47169
QY 4743 tgcgaatggcaagcccacatcaatggcaccgatggcaagcccccaaaagccatcaaggcgc 4802
Db 47170 ----- 47169
QY 4803 cgaatgtaataatactacgcccacacgcaacgcgtacctgtggacaagaatggcaagcc 4862
Db 47170 ----- 47169
QY 4863 catcaccgatgggacaacacttgccaaacttggaagctcattggcaaaccttgcagcag 4922
Db 47170 ----- 47169
QY 4923 tcaatcaagtgtgtggaagcctagcggcaactcagatgcacatcaccttaaccaacatcaa 4982
Db 47170 ----- 47169
QY 4983 gtccactttgccaaaatlgacacacacacaggttaatggccaatgcagggcaagccca 5042
Db 47170 ----- 47169
QY 5043 aagtcgccagcctatcagcagcagcaagaaatgctgcccagtgtaaaagtgtgct 5102
Db 47170 ----- 47169
QY 5103 aaatgtaggctttaacttgacagccaatcaaatcaatgaactttgtcaaaagcctatga 5162
Db 47170 ----- 47169
QY 5163 taccgtcaactttgtcaatggtacaggtgcgcacatcacaaagcgtgcgtagtgctgag 5222
Db 47170 ----- 47169
QY 5223 cagcgtagtaacatcacccgtcaacaccgccttagcagcagcagatgatggcaatgt 5282
Db 47170 ----- 47169
QY 5283 gcttatcaaaagcaaaagatggttaagttctacaaagcagacgaacctcatgccaaacggctc 5342
Db 47170 ----- 47169
QY 5343 actaaaagcaggcaaatcagccagtgatgccaaaactccaactggttaagccttgttaa 5402
Db 47170 ----- 47169
QY 5403 ccccaatgctgtaaaagcagttacagggatgcagtggtgctttaataacttatcaaaagc 5462
Db 47170 ----- 47169
QY 5463 ggtatttaaatccaaagatggttaacactactacacagtaagcctctgatggcatcagtat 5522
Db 47170 ----- 47169
QY 5523 ccaagggcaagaataacagcagcatcaccttaagcaaaagatgggctgaatgtaggcggtaa 5582
Db 47170 ----- 47171
QY 5583 ggtcactagcaatgtggttaaaagcacaagacacccagcgtgcgaatgtacaacagtt 5642
Db 47172 GGTCAATCAGCAATGTGGGTAAAGGCACAAAAGACACCGACGCTGCCAATGTACAAACAGTT 47231
QY 5643 aaacggaagtgcgaactgttgggtcttggtaatgctggttaataagataaacgctgacggcaa 5702
|||||

Db 47232 AAACGAAGTACGCAACTTGTGGGTCTTGGTAATGCTGGTAAATGATAAACGCTGACGGCAA 47291
Qy 5703 tcaagtaaacattgcccgcacatacaaaaagaccacaaattcaggtttcatcatcaataccgac 5762
Db 47292 TCAGGTAAACATTGCGGACATCAAAAAGACCCAAATTCAGTTTCATCATTAACCCGAC 47351
Qy 5763 tgcataaagcaggacaggtacttggcgttaaggttaaggttaataaacgataccgaaacacttgc 5822
Db 47352 TGTATCAAAAGCAGCAGCGTACTTGGCGGTAAAGGTAAATAACGATACCCGAAACACTTGC 47411
Qy 5823 caactggtgtgataagtggtgataaagagcgcgaacgctaacgagcgtatttaagcaa 5882
Db 47412 CACTGTGTGTATACAGTGGCGTGGTAAAGACGCGCAACGCTAACGGCGATTTAAGCAA 47471
Qy 5883 tgtttgggtcaaaaccccaaaagatggcagcaaaaaagccctgctgcacattataacgc 5942
Db 47472 TGTGTGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCACATTATAACGC 47531
Qy 5943 cgcaggtcaagacaactatttgacacaacacccgcagaagccattgacagaataaataga 6002
Db 47532 CGCAGGTCAGCAACCACTATTGACCAACAACCCCGCAGAACCCATTGACAGAAATAGA 47591
Qy 6003 aceagtatccgtttcttcattgcaacgatggcaatcaagagcctgtgtacaaggcg 6062
Db 47592 ACNAGGTATCCGCTCTTCCATGTCAACGATGCAATCAAGAGCCGTGGTACAAAGGCG 47651
Qy 6063 taacgcatgactcaagtctcaggaagcaactcagtggtgagatagtttccagggcaa 6122
Db 47652 TAACGGCATTTGATTTCAAGTCTGCTGGCAAGCATTCAGTGGCGATAGGTTTCCAGGCCAA 47711
Qy 6123 ggcagatggtgaaagccgcgttgcctataggcagacaaaccccaagcaggcaaccaatccat 6182
Db 47712 GGCAGATGTTGAAGCCGCGTGTCCATAGGCAGACAAACCCAGCAGGCAATCCAT 47771
Qy 6183 cgcagatggtgataacgcagaaagccagggcgatcaatcatcgccatcggttacaggcaa 6242
Db 47772 CGCCATCGGTGATTAAGCAACCAAGCAGCGGCGATCAATCATCGCCATCGGTACAGGCAA 47831
Qy 6243 tftggttaacaggttaagcaactctgtgcatcgccagcccaagcaactgtttaaggtgataa 6302
Db 47832 TGTGTTAGCAGGTAAAGCACTCTGTGTCATCGCGACCCCAAGCACTGTAAAGCCTGATAA 47891
Qy 6303 cagttacagtggtggttaataaacacagtttattcgatgccactcaacagatctttgg 6362
Db 47892 CAGTTACAGTGTGGTGAATTAACAACCAAGTTTACCAGATGCCACTTCAGCCGATGCTTTGG 47951
Qy 6363 tgtgggcaataacataccctgaccgaaagttaactcgtggttgccttgggttcaaacctcgc 6422
Db 47952 TGTGGGCAATAACATACCGTGACCGAAAGTAACCTCGTTGCTTAGGTTCAAACTCTGC 48011
Qy 6423 catcagtcagggcacacacgcaggacacacagcccaaaatctgacggcacacaggtac 6482
Db 48012 CATCAGTCGAGGCACACACGACGACACAAAGCCAAAATCTGACGGGCACACAGGTAC 48071
Qy 6483 aaccacacacaggtgcaacaggtacggttaaggtttgctggacaaacgcggttgg 6542
Db 48072 AACCACACACAGCTGCCACAGTACGGTTAAAGGCTTGTCTGGACAAACGCGGTTGG 48131
Qy 6543 tgcggtctccgtgggtgctcaggtgtgtgaacgcggtatccaaaaatgtgycagcaggtga 6602
Db 48132 TGCAGTCTCCGTGGGTGCTCAGGTGTGAACGCGGTATCCAAAATGTGCGACGAGTGA 48191
Qy 6603 ggtcagttcaacacagcaggtatcggtcaatggttagccagttgtacaaagccacccaag 6662
Db 48192 GGTCACTGCCACACAGCAGCGATGGCGTCAATGTTAGCCAGTGTGTACAAAGCCCAAG 48251
Qy 6663 cattgccaacgcaaccaaagcttgaccatcgatccacacaaacgaaataaagccaa 6722
Db 48252 CATTCGCCAACGCAACCAATGAGTTGACCATGTCATCCACAAAACGAAATAAAGCCAA 48311
Qy 6723 tgcagggatttcaatcagcagtggtgagtggtccatgcccaagcctacattcctggcag 6782
Db 48312 TGCAGGGATTTTCATCAGCGATGGCGATGGCGTCCATGTCACNAGCCTACATTCCTCGCAG 48371

Qy 6783 atccatggttacgggggtattgtccaccacacacaggtcaaggtgcggtggtgagtgact 6842
Db 48372 ATCCATGTTTACCGGGGTATTGCCACACCCACACAGCGTCAAGGTGCGTGGCAGTGGGACT 48431
Qy 6843 gtcqaagctgtcgtgataaattggtcaatgggttatttaaaaaatcaatggttcagcgataccaca 6902
Db 48432 GTCGAAGCTGCGGATAATGGTCAATGGGTATTATAAATCAATGTTTCAGCCGATACCCA 48491
Qy 6903 aggcattgtaggggcgcgaggtggtggtcaggttttcaacttt 6942
Db 48492 AGCCATGTAGGGGACGACGTTGGTGCAGGTTTTCACTTT 48531
RESULT 4
AX079918 AX079918 6141 bp DNA PAT 22-FEB-2001
LOCUS Sequence 6 from Patent WO0107619.
DEFINITION AX079918
ACCESSION AX079918
VERSION AX079918.1 GI:13159439
KEYWORDS Moraxella catarrhalis.
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
REFERENCE 1 (bases 1 to 6141)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of moraxella
JOURNAL Patent: WO 0107619-A 6 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
source Location/Qualifiers
1. 6141
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 2012 a 1403 c 1399 g 1327 t
ORIGIN
Query Match 37.3%; Score 2590; DB 10; Length 6141;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 4802; Conservative 0; Mismatches 1105; Indels 1269; Gaps 28;
Qy 1 atgaatcacatataaaagtcatctttaacaaagccacagggcacatttatggcgtggca 60
Db 1 ATGAATCACATCTATAAAGTCATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGGCA 60
Qy 61 gaggcgcacaaatccacacagggagggagtagcagtagcagcagcagcagcagcagcagc 120
Db 61 GAGTACGCCAAATCCACAGACGCGGGGGGGTACTGTGTACAGGGCAGTTGGCAGT 120
Qy 121 tctcctgtcatccgctgactcgtgtgtccacgctcgctatactctagtgtgctgagcag 180
Db 121 GTATGCACCTCTGAGCTTTGCCCGTATTGGCGGCTCGCTGTCTGTGATCGTGCAACG 180
Qy 181 ctcaatggcagtgcttatgctcaaaa-----ataatagcaagatcgacttgg----- 228
Db 181 CTCAGTGGCAGTGCTTATGCTCAAAAAAAGATACCAACATATCGCAATTGGTGAACAA 240
Qy 229 ----accacagggcaacaatgacaatgcctcggtgtagcaatgaagcagcatcctattggt 285
Db 241 AACCAGCCAAAGCGCTCAGGCATCGCAAGCGGAGGGGTATCGAGCCATTGCTATTGTT 300
Qy 286 agtctgttaaggcacatgccaatcaagctattgtctatcggtggtagcaaacagatcct 345
Db 301 GAAAAATGCTAACACACAGGCGGTCGAAGCATCGCCATCGGTAGTAGTAA---TAAACT 357
Qy 346 cgtaatcaagcggcctaatacagaagcaggttcccacgccaaggttaagagtcacacgc 405
Db 358 GTCAATGGGAAGCAGTTTGGATACATAGGTACCGATGCTACGGGTCAAGAGTCCATCGCC 417
Qy 406 atcgggtggtgagtactggtggtggtggtggtggtggtggtggtggtggtggtggtggt 465

Db 418 ATCGGTGGTGTATTAAGGCTAGTGGTGTGATGCCCTCGATTCGCCATCGGTAGTGTAGTACTTA 477
QY 466 tatttgatgaatagcactaactcttaataatccaaatgg---tcttcttagcactctt 522
Db 478 CAITTTGCTTATGATGATGATGATTAATTAACATCCGAAGGTACTCTGATTAACGATCTT 537
QY 523 attcaaaaccatacagattattacgccaaataacgagactcaaatggttctca---gaaatat 579
Db 538 ATTAACGGCCATCGATTTTAAAGAAATACGAAGCTCAAGAGTAAATGATGTAATAATAT 597
QY 580 agagcacagcagcagaagacacgcagctactgctgagtgagggccatggccatgcaatgcaaaag 639
Db 598 AGACGCACACCCGACAGCCGACAGCCAGTACTGTCAGTGGAGCCCATGCTATATGACAG 657
QY 640 ggtcattttcccaacgctttggtacacggtcaccagctgaaagcaactatctcttgcca 699
Db 658 GGTCAITTTTCCAAAGCCTTTGGTACACGGGCAACAGCTAAAAGTGCCTATTCTCTTGGCA 717
QY 700 gtaggcttaccgccaagcgaaggaaggaatatacaaatcgctattggttcttaataatgcaaa 759
Db 718 GTGGGTCTTCCGCCACAGCCGAGGGCCAATCTACAATCGCTATTGGTCTGTGATGCAACA 777
QY 760 gctatcaattatgagcactagcccttgggtgcagatactcagtggtgatttggattacggt 819
Db 778 TCTAGTCTGTTGGAGCGCATAGCCCTTGGTGCAGGTACTCTGCTCAGCTACAGGCGAGT 837
QY 820 attgcccattggttattgctcagatccttaataataataataata---ataataaa 876
Db 838 ATTGCCCTAGGTCAAGGTTCTGTTGTCACCTCAGAGTGATATAATTTCTAGACCGGCTAT 897
QY 877 gctatgtaccgaaggaatggtgtaacacataaaatcgctcctaaagccacgcgcaatggt 936
Db 898 ACACCAANTACCCGAGCCTAGACCCCAAGTTTCAAGCCACCAATATACGAAGGCGGGT 957
QY 937 ttatttccattgtagtagcactatacaagctcaaatcaatcaatgctcgtgaggttat 996
Db 958 CCATTTCATTTGGTAGTAATCTATCAACAGTAAATCATCATATGTCGTGTCAGGTGTT 1017
QY 997 gaggataccgctggttcaatgtggcacagctaaagcgtggagaaatctggcta---ag 1053
Db 1018 AATAAACCAGTACGGTCAATGTGGCAGCTAGAACGGTGGTGAAGTGGGCTAAGGAG 1077
QY 1054 cgtcaattactttaaaggtgatgataacggtactggttgaagaaataactggcgag 1113
Db 1078 CGTAGAATTAATTTACGGGTGATGATAACAGTACTGACGTAATAATAGTTTGGATAAT 1137
QY 1114 acttcaaccattaaaggtgtgagaccgaagcgacaagctaaaccgataataataacatt 1173
Db 1138 ACTTTAACTATTAAAGTGGTG---CAGAGACCAACGCATTAAACGATATAATATCGGT 1194
QY 1174 ggtgtgtaacagataataacttgctgaaagttaaacttgctaaaaacctaagcgtt 1233
Db 1195 GTGTAAAGAGGCTGATAATAGTGGTCTGAAAGTTAAACTTGTCTAAACCTTTAAACAAT 1254
QY 1234 ctggaacaggttagcaccacaaacctaacccgagcgagagaagttacgtag---tgagt 1290
Db 1255 CTACTGAGGTGATACACTACATTAATGCCCACACACAGTAAAGGTAGGTAGTAGT 1314
QY 1291 ggttaataaccgctgagctacaaaagcgttgggtttaaacttttacc---caacaaca 1344
Db 1315 AGTAGTACTACAGCTGAATTTATTGAGTGTATGATTTAACTTTACCCAGCCCAATACAGGC 1374
QY 1345 aatgcaagcagacacaaaacccgtctatggaactgaggttgaagtttactgataatct 1404
Db 1375 AGTCAAGACCAAGCAAAACCGTCTATGGGCTTAATGGGTGAAGTTTACTTAATAATGCA 1434
QY 1405 aatcggcacttga---gatactactcgtatcaccaagaataaattggttttagcaat 1461
Db 1435 GAACAACACAGCAATTCGGCACTACTCGTATTACCAGAGATAAAATTTGGCTTTGCTCGA 1494
QY 1462 aaagctggtcagttgataaacaaccttattctgtataaagacaaagctaaaaag---- 1516
Db 1495 GATGGTGTATGTTATGATGAACCAAGCACCATATTGATGAACCAACCACTTAAAGTGGGT 1554

QY 1517 -----tt 1518
Db 1555 AGTGTTCGAATTACCATAGACAATGGCATTTGATCGAGGTAAATAAAAAGATCAGTAATCTT 1614
QY 1519 ggaacaagcaccctaaacaacaggtggttactgttactgttaataacaccattggtgtagcaat 1578
Db 1615 GCCAAAGGTAGCAGTGTAAAGCTGCGGTACCATCAACAGCTCAAAAGCGCCAAAGCCT 1674
QY 1579 aaacaataccaagtcggtgctgatggcattaaaatttcccgatgtgaatgttaagtatca 1638
Db 1675 ACITTTAAACCCAGCGCTGCGATCAGTGTACACCTACTGAAATATCAGTTGATGCTTAAG 1734
QY 1639 aatgcgcgaataatcggcactactcgtattaccogaagagaaattggtcttgcgtgact 1698
Db 1735 AGTGGCAATTTACGCCCCCAACTTACACATTTGGCTGAAACACCCAGGAGCTTAACAGT 1794
QY 1699 gatggtaaagtgtataa----- 1715
Db 1795 GATGGCACTTAGTGTCAATAATTTAGTGTAAAGGTAGTGGTACGAACAATAGCTTTAGTTACC 1854
QY 1716 -----aaagtcaccatatttgataaaaaaacaacttcaagtggtggtt 1761
Db 1855 GCCGAACATTTGGCAACCTATCTAAATGAAGTCAATCGAACGGCTGACAGTCTTACAA 1914
QY 1762 aaattaccaaaacagtggtcattaatgacggtgatcaaaagatcagtaattgttaagat 1821
Db 1915 AGCTTTACCGTTAAAGAAGAAGACGATGACGCCAACGCTATCACCGTGGCTAAAGAT 1974
QY 1822 gcaacgagcagatccgagtcagtcactataaaacaggttaaaacagtcacaagaagcgc 1881
Db 1975 ACGACAAAAAATCGCGGCGAGTCAGCATCTTAAATCTCAAAAGTAAACCGGTCTAACG 2034
QY 1882 gacggtgcccataaa-----agcttctcttctcgtgatgaaaggttcagaa 1929
Db 2035 GTTGTACCAAAAAAGATGTTACGGTTACCTTTGGGGTGTAGCCAGATAGCGGTCTGACC 2094
QY 1930 ttacgattagtaacttgtattcttaattggttaataaccocaaatacctttgagacatcacc 1989
Db 2095 ATTGGCAAAAGCACCCCTAAACACAGCATGGCTGACTGTTAAAGATACCAACGAACAATC 2154
QY 1990 ttgcaagtgaaaacggtcagtcagtcagcaatgacatagccaagaagtaaaagtcacaaag-- 2047
Db 2155 CAAGTCGGCTGTAATGGCGATTAATTTACTAATGTGAATGTGTAGTAATCCAGGTACTGGC 2214
QY 2048 -ttggtattgaccaatcaatggctcaccacgctcaagctgacctgggtgtagcgataaa 2106
Db 2215 ATTGCAAAATACCGCTCCGATTTACCAGAGATAAAATTTGGCTTGGTCTGTGATGGTGCA 2274
QY 2107 gatggttaaaactcaattggttattgagcaag----- 2137
Db 2275 GTTGATACAAACAACCTTATCTTTGATCAAGACAAGCTACAAGTTGGCAATGTTAAGATT 2334
QY 2138 ----tggtacggtgtaacgacacacaaatacattatagaggtgtgtcccaacactgcct 2193
Db 2335 ACCAACACTGGCATTAACAGGTTGTTAAAGCCATACAGGGCTGTCCCCAACACTGGCT 2394
QY 2194 agcattaccatgaggtggtgctacgcacacacagagggcaatacaataccaccagcag 2253
Db 2395 AGCATTCGGCATCAAGTAGC---CGCAACATACAACTAGGCAATACAAATCCA---AGAC 2448
QY 2254 gaagcaaaatccaaagcgcagctacgttgcgtatataataatacagcgttttaacctaaaa 2313
Db 2449 AAAGACAAATCCAAACGCTGCCAGCATTAATGATATATAATACAGGCTTTAAACCTTAAA 2508
QY 2314 aataatgacactcgtggtgcttctcacttataaacacttaaacacttaacactgagc 2373
Db 2509 AATAATAACAAACCCCATTTGACTTTGTCTCCACTTATGACATTTGTGTGCTTTGCCAATGGC 2568
QY 2374 aatgccaccacggttaagtaacttacgatgaacacaaatacaacagtaagtaacttat 2433
Db 2569 AATGCCACCCGCCACAGTAACCCATGATACCCGCTAACAAAAACCACTTAAAGTGTATAT 2628

QY	2434	gatgtcaatgttgatgagaaaaacacattgaactcaaggcgataataatggcagacacaaaca	2493
Db	2629	GATGTCAATGTGGATGATACAACCATTTCACTTAACAGGCACATGATGACAA---TAAAAAA	2685
QY	2494	attggcgtcaaaacacacacactgacacacaaatgtctaatgggtaaagcaacc-----	2547
Db	2686	CTTGGCGTCANAAACCNCNACTGMAACAAACAAGTGCTAATGGTAAATACAGCAACTAAC	2745
QY	2548	---aactttagtaacacogataacogatgccttgtttaacgcgcaaaagacatcgccgaat	2604
Db	2746	TTTAAATGTTAACTCTAGTATGAAGATGCGCTTTGTTAAAGCCCAAGACATGCCGAAT	2805
QY	2605	ctaaacacctagccaagaaattcacacacaaaggcacagcagacacgcgcctacaa	2664
Db	2806	CTAAACACCTTAGCCACAGAAATTCACACCACAAAGGCACAGCAGACACGCCCTACAA	2865
QY	2665	accttaagttcaaaaagacggtgca-----actgatgacgaaaacccatcacgtgggt	2718
Db	2866	ACCTTTACCGTTTAAAGAGGTAGATGNAATTAATATGCTGATGACGCCAACGCCATCAC	2925
QY	2719	aaagatggttacaaaaacgcgcgaagccgtcaacactctaaactcaaaagtgaaaaacggt	2778
Db	2926	GTGGTCAAAAAGACGCAAAATAATCAAGTCAACACGCTTAACACTCAAAAGTGAAAAACGT	2985
QY	2779	ctaacggttctaccaataaagatggttacggttacctttggtatataaacccaagcgt	2838
Db	2986	CTTAAATATTAAACACGCAAAAATGTTACGGTTACCTTTGGCATTTAACACCACAGCGGT	3045
QY	2839	cttaagccgcgcagcaccactctaaacaaagatgctgtctattataaaaccccgct	2898
Db	3046	CTTAAAGCCGGCANAAAGCACCC---CTAAACGACGGTGGCTTGTCTATTAAAAACCCCACT	3102
QY	2899	agtaacgaaacaaatccaagtcggtgctgtagcgtgaagtttccaaagt--tgataag	2955
Db	3103	GGTAGCCAAACAAATCCAAGTCCGGTCTCATGGCTGAAGTTTGGCAAGGTTAATAATAAT	3162
QY	2956	ggttaattcaagcactggctttagtgcaacagcgtatcaccaagaatcataattggcttt	3015
Db	3163	GGTGTGTAGTCTGGCATTCATGSCAACACTCGCATTAACAGAGATGAATTTGGCTTT	3222
QY	3016	actggggctaatggctcaottgataccacccaaacccacccaataaccaagaagcttaa	3075
Db	3223	ACTGGGACTAATGGCTACCTTGATAAAGCAAAACCCCACTTAGCAAGAC-----	3273
QY	3076	gtgggtgaagttgaaattaccaacactggcattcaacgcaggtggtgtaaaaagattaccaac	3135
Db	3274	-----GSCATTAACGCAGGTGTTAAAAAGATTACCAAC	3306
QY	3136	attcaatcaggtgatattaccaaaacagcaatgatgctgtgacagcgctcggtttat	3195
Db	3307	ATTCAATCAGGTGAGATTGCCAAAAACGCCATGATGCTGTGACAGGGCGCAAGATTTAT	3366
QY	3196	gatttaaaacccgaacttgaagcaaaaatcaacagctgctgctaaaaacagcacaaaaactoa	3255
Db	3367	GATTTAAANCCGANCTTGAANACAAATCAGCAGTACTGCCAAAACAGCACAAAACCTCA	3426
QY	3256	ttacacgaattctcagtagcagatgaacaggtatcaactttagtcaggttagtaacccttac	3315
Db	3427	TTACACGAATTTCTCAGTAGCAGATGAACAAGGTAATAACTTTACGGTTAGTTAACCCCTTAC	3486
QY	3316	tccagttatgacacctcaagacctctgatgcatcacctttgcaggtgaaacacggcatt	3375
Db	3487	TCGAGTTATGACACTCAAGACCTCTGATGTCATCACCTTTGCAGGTTGAAACGGCATT	3546
QY	3376	accaccaaggtataaaggtgtgtgtgcgtgtggtgcattgacccaaacaaagggcttaacc	3435
Db	3547	ACCACCAAGGTAAATGAAGGTGTGGTGTGGGATTTGACCAAAACCAAAAGGGTTAACC	3606
QY	3436	agcctaagctgacccgttggtataataataatggcaaaagcattgctcattgacagtaagat	3495
Db	3607	AGCCCTTAGCTGACCGTGGGTAAATAATATGCGCAAGGCATTTGTCATTGACAGCCAAAT	3666
QY	3496	ggtcaaaatcaccatcacaggactaagcaacactctagctaaatggttaccaatga-----t	3549

Db	3667	GGTCAAAATACCATCAGGACTAAGCAACACTTGTAGCTANTGTTTACCATGATAAGGT	3726
Qy	3550	ggtcagagacgcacctaagccaaggcttgccaatgcacaccgacaaaacccgtgcgcgc	3609
Db	3727	AGCGTACCACACACAGAACAGGGCAATATATCAAGACAGCAACAAACCCGCTGCCGC	3786
Qy	3610	agcattggtgatgtgtctaaacgcagggttttaacttgcaggccaagtgtgaagcgggtgac	3669
Db	3787	AGCATTTGTTGATGTGCTAAGCGCAGGCTTTAACTTGCAGGCAATGGTGAAGCGGTGAC	3846
Qy	3670	tttgtctcaacttatgacactgttgactttatcgatggcaatgcaccacccgttaagggtg	3729
Db	3847	TTTGTGTCACACTTATGACACCGTCACTTTGCCGATGGCAATGCCACCCGCTAAGGTG	3906
Qy	3730	acctatgatgacacgaacaaacccagtaaaagtgtgtctatgatgtcaatgttgataataaa	3789
Db	3907	ACCTATGATGACACAGCAACCAACAGTAAAGTGGTCTATGATGTCAATGTGGATGATACA	3966
Qy	3790	accattgaagtgaacagtgataaaaaacttggcgtcmetaaacccacacacagtcacaaaca	3849
Db	3967	ACCATTTGAATTTAAA--GATAAAAAACTTGGCGTAAAAACCACCACTTGACCACTACT	4023
Qy	3850	agtgcataagttaagtca---accataattagtcgcgcgcatggcgatgccttctgttaa	3906
Db	4024	GGCACAGGTGCTAATAAATTTGCCCTAAGCAATCAAGCTACTGGCGATGGCTTTGCAAG	4083
Qy	3907	gccagtgaatgcgcaccacatataacaccttggctgcgcgacatccaaacgcgcaagg	3966
Db	4084	GCCAGTGATATCGTTGCTCATCTAAACACCTTATCTGGCGACATCCAACTGCCAAGGG	4143
Qy	3967	gcaagccaaagcagctccagcaagctatgttgatgtgatggcaacaaggttcatttat	4026
Db	4144	GCAAGCCACCGAACAACTCAGACGCTATGTGGATGCTGATGGCAATTAAGTGTATCTAT	4203
Qy	4027	gacgttacogataagaagtactatacagtcacatgacaagaggtcgaatgacaaaacaaa	4086
Db	4204	GACAGTACCGATTAACAAGTACTATCAAGCGAAAAATGATGGCAGTGTATMAACCAAA	4263
Qy	4087	gaagtgccaaagacaaaactggtcgccaaagccaaaccccgatggcacattgggtcaa	4146
Db	4264	GAGTTGCCAAAGCAAACTGGTGCGCCAAGCCAAACCCAGATGGCAATTTGCTCNA	4323
Qy	4147	atgaatgtcaaatcagtcattacaagaagcgaagtaaatgtagtccaatataaaagcaagc	4206
Db	4324	ATGAATGTCAATCAGTCATTAACAAGAACAAAGTAAATGATGCCAATAAAAAGCAGGC	4383
Qy	4207	atcaatgaacacacgccttatacgaagggcttgaaaaacgcgcgcaaaagacaccaaac	4266
Db	4384	ATCAATGAAGACAAACGCTTTGTTAAAGNACTTGAANAAGCCGCTTCTGATAACAAACC	4443
Qy	4267	aaaacgcgcagtaactgtgggtattaaatgcggttgccaaacacgcgtgacctt	4326
Db	4444	ANAAACGCCGAGTAACGTGTGGGTGATTTAAATGCGGTTGCCCAACACCGTGCCTTT	4503
Qy	4327	gcagggatatacaggcacaacgcgtataaaactggcgagactttgaccttgcataaagtg	4386
Db	4504	GCAGGGATACAGCACACCGGCTTAAAAAATCTGGCGAGACTTTTGACCATCAAGGTGG	4563
Qy	4387	caaacagacaccaataaagctaaacgataataacatcgggtgtgtagcaggtactgagtc	4446
Db	4564	CAACAGACACCAATAGCTTAACCGATATAATACATCGGTGTGTAGCAGTACTGATGGC	4623
Qy	4447	ttcactgtcaacttgcaagacctaacaatttaacagcgttaagtgcaggtgcacc	4506
Db	4624	TTCAGTGTCAACTTGCACAAAGACCTTAACCAATCTTAACAGCGTTAATGCAGGTGGCACC	4683
Qy	4507	agaaatgatgataaaaggcatctcttttgtagacgcaaacggttcaagccaaagcaaacacc	4566
Db	4684	AAAAATTGATGACAAAGGCGCTGCTTTTGTAGACTCAAGCGTCAAGCCAAAGCAACACC	4743
Qy	4567	cctgtgctaagtgcacaaatgggctggaaactgggtgcgaacgcatacagtaacatcggtgca	4626

Db 4744 CCTGTGCTAAGTGCATATGGCTGGACCTTGGGTGGC----- 4779
Qy 4627 gctgtgatgataacgagtgcggtgaactttaagcagtttaagagtggtgcacaaacggtc 4686
Db 4780 ----- 4779
Qy 4687 acaacctaacaacaaagtaactcaggtgctgctatcccttttggttaacogatgcc 4746
Db 4780 ----- 4779
Qy 4747 aatggcaagcccatcaatgacacgagtggaagcccccaaaagccatcaagggcgccgat 4806
Db 4780 ----- 4779
Qy 4807 ggttaactatcacgccaacgccaacgctgacacgctgacaaagatggcaagcccatc 4866
Db 4780 ----- 4779
Qy 4867 accgatcggaacaaacttgccaatctggcagctcatggcaaaccccttgatgacaggtcat 4926
Db 4780 ----- 4779
Qy 4927 caagtgtggaagcctagggcgaactcagatgccatcacctcaacaaacatcaagtcc 4986
Db 4780 ----- 4779
Qy 4987 actttgccaaaattgacacacacaaacacaggttaagtccaatgagggcaagcccaagt 5046
Db 4780 ----- 4779
Qy 5047 ctgcccagcctatcagcagcacagcaaaagtaactgtccaggtgtcaaaagatgctgaat 5106
Db 4780 ----- 4779
Qy 5107 gtaggctttaacttgcagacaaatcacaatcaagtggactttgtcmaaagcctatgatacc 5166
Db 4780 ----- 4779
Qy 5167 gtcaacttgtcaatggataggtgcgcgacatcaacaagcgtgcgtatgctgatggcacg 5226
Db 4780 ----- 4779
Qy 5227 atgagtaacatcacccgtcaacacccgcttagcagcagccgatgatgagtgcaatgctctt 5286
Db 4780 ----- 4779
Qy 5287 atcaaaagcaaatggttaagttctacaagcagacacctcatgccaaacggctcacta 5346
Db 4780 ----- 4779
Qy 5347 aaagcaggcaaatcagccagtgatgccaacaaactccaactggtctaaagccttgttaacccc 5406
Db 4780 ----- 4779
Qy 5407 aatgctgtaaaagcagtagcaggtgagtgctctttaaacttatcaaaagcaggtta 5466
Db 4780 ----- 4779
Qy 5467 tttaaatccaaagatggttacaactactaccacagtaagctctgatggcatcatccaa 5526
Db 4780 ----- 4779
Qy 5527 ggcaaaagataacagcagcatcaccctcaagcaaatggctgaatgtaggcggtaaggtc 5586
Db 4780 -----AAGGTC 4785
Qy 5587 atcagaatgtggttaaaagcacaacaaagacacccagcctgccaatgtacaaacagcttaaac 5646
Db 4786 ATCAGTAATGTGGCAAGGCACAAAAGATACCGACGCTGCCAATGTACAAACAGATTAAAC 4845
Qy 5647 gaagtacgaactgttgggtctgtggttaagtctggttaataacgctgacggcgaatcag 5706
Db 4846 GAAGTAGCAACTTGTGGGTCTTGGTAAATGCTGTAATGATACCGCTGACGGCAATCAG 4905

Qy 5707 gtaaacattccgacatcaaaaaagaccccaattcaagttcatctatctaaacgcactgtc 5766
Db 4906 GTRAACATTCCGGACATCAAAAAGACCCAAATTCAGTTCATCATCTAACCGCACGTGTC 4965
Qy 5767 atcaaaagcagcaggtacttgccgttaaaaggttaataacgataccgaaaaacttgcacat 5826
Db 4966 ATCAAAAGCAGCAGGTACTTGGCGGTAAGAGTAATTAACGATACCGAAAACTTGCACCT 5025
Qy 5827 ggtggtgtacaagtggcggtggtataaagacggaacgctaagcgtacgagcgttaagcaatgtt 5886
Db 5026 GGTGGTATACAGTGGCGGTGGATAAAGACGGCAACGCTTACGGCGATTAAAGCAATGTT 5085
Qy 5887 tgggtcaaaaccccaaaagatggcagcaaaaaagccctgctgcaccttataacgcgcga 5946
Db 5086 TGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTGCCACTTATAACGCCGCA 5145
Qy 5947 ggtcagaccaaactatttgacaaacaaacccgcgagaagccattgacagaataataatgaacaa 6006
Db 5146 GGTGAGACCAACTATTGTGACCAACACCCCGCAGAAGCCATTGACAGATAAATGAACAA 5205
Qy 6007 ggtatccgcttctccatgtcaacgatggcaatcaagagcctgtgtacaaagcgtaac 6066
Db 5206 GGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGGTACAAAGGCGTAAC 5265
Qy 6067 ggcattgactcaagtgcctcaggcgaagcactcagtgcgataggtttccagggccaaagca 6126
Db 5266 GGCATTGACTCAAGTGCCTCAGGCAAGCACTCACTGGCGATAGTTCAGGCCCAAGGCA 5325
Qy 6127 gatgtgaagcgcgcttgccatagggcagacaaaccccaagcaggcaacccatccatcgcc 6186
Db 5326 GATGGTGAAGCGCGCTTGGCCATAGGCAGACAAACCCCAAGCAGCAACCAATCATCGCC 5385
Qy 6187 atcgtgtatacgcacaaagccacggcgatcaatccatcgcatcggtacagcgaatgtg 6246
Db 5386 ATCGGTATACGCAACGACCCGCGGATCAATCCATCGCCATCGGTACAGGCAATGTG 5445
Qy 6247 gtaacaggtgaagcactctgtgccatcgcgaccccaagcactgttaaggctgataacagt 6306
Db 5446 GTAGCAGGTAAAGCACTCTGTGGTCCATCGGCAGCCCAAGCACTGTAAAGCTGATAACAGT 5505
Qy 6307 tacaagtgggtaataacaacacagtttatcgatgccactcagacccgatgtctttgggtg 6366
Db 5506 TACAGTGGGTATAACAACCACTTTACCGATGCCACTCAAAACCGATGTCTTTGGGTGTG 5565
Qy 6367 ggcaataacatcacctgacgcgaagtaactcgtgtgccttaggttcaaaactctgcacatc 6426
Db 5566 GGCATATACATCACCGTGCACGAAAGTAACTCGTTGCCCTTAGGTTCAAACTCTGCCATC 5625
Qy 6427 agtcagggcacacacgcagcagcacacaagcccaaaaaatctgacggcacagcaggttacaacc 6486
Db 5626 AGTCAGGCACACACGCGAGGCACACAAGCCAAAAAATCTGACGGCACAGCAGGTACAACC 5685
Qy 6487 accacagaggtgcaacaggtacggttaaaagctttctggacaaaacggcggttgggtgcg 6546
Db 5686 ACCACAGAGGTGCAACCGGTACCGTTAAAGGCTTTGCTGGACAAACGGCGGTGGGTGCG 5745
Qy 6547 gtcctcgggtgctcaggtgctgaacgcgctatcaaaaaatgtggcagcaggtgaggtc 6606
Db 5746 GTCTCCGTGGGTGCTCAGGTGCTGACCGCGTATCCAAAATGTGGCAGCAGGTGAGGTC 5805
Qy 6607 agtgcaccagcagcatggggtcaatggtagccagttgtacaaagccaccccaagggcatt 6666
Db 5806 AGTGCCACACGACCGCATGCGGTCAATGGTAGCCAGTTGTACAAAAGCCACCAAGCATTT 5865
Qy 6667 gcaaacgcaacaaatagccttgaccatcgatccacacaaacgaaaaataaaagcacaatgca 6726
Db 5866 GCCAACGCAACCAATAGCTTGACCATGTATCCACAAAACGAAAAATAAGGGCCAAATGCA 5925
Qy 6727 gggatttcacagcagtggtggtcgtccatgccaaagcctcatcttctcgtgagatcc 6786
Db 5926 GGGATTTCATCAGCGATGGCGATGGCGTCCATGCCAANAGCCTACATTCCTGGCAGATCC 5985

Db 4208 GGTCAAAATACCATCACAGGACTAAGCAACACTCTAGCTAATGTTACCAATGATAAAGGT 4267
 QY 3550 ggtcaggacacgcactaagccaagggtctgccaatgacacccgacaaaacccgtgcgcgc 3609
 Db 4268 AGCGTACGCACACAGACAGGCGAATAATATCAAGACGAGACAAACCCGTCGCCGCC 4327
 QY 3610 agcattggtgatgtctaaacgcagcgtttaaactgcaagcaatggtgaagcgggtgac 3669
 Db 4328 AGCATGTTGATGTGCTAAGCGCAGCGTTTAACTTGCAGGCAATGGTGAAGCGGTTGAC 4387
 QY 3670 ttgtctccactatgacactgttgactttatcgatggcaatgccaccacgcgttaagtg 3729
 Db 4388 TTTGTCTCCACTTATGACACCGTCAACTTTGCCGATGCAATGGCCACCCACCGTAAAGTG 4447
 QY 3730 acctatgatgacacgaagcaaaacccagtgtaagtggtctctatgatgcaatgtggataataaa 3789
 Db 4448 ACCTATGATGACACGAAGCAAAACCCAGTAGTGGTCTATGATGTGATGATGATACA 4507
 QY 3790 accttgaagtacagtgataaaaacttggctcaaaacccacacacactgacccaataa 3849
 Db 4508 ACCATGAGAGTTAA--GATAAAAAACTTGGCGTAAAAACCCACCATTGACCAAGTACT 4564
 QY 3850 agtgcataatggttaatgca---accataatagtcgcgcgcgtgagtgcccttgttaa 3906
 Db 4565 GGCACAGGTGCTAATAAATTTGCCCTAAGCAATCAAGCTACTGGCGATCGCGTTCTCAAG 4624
 QY 3907 gccagtgtatgccaccatcaataacttggctggcgacatccaaacccgcaaaagg 3966
 Db 4625 GCAGTGATATGTTGCTCATCTAAACACCTTATCTGGGAGATCCAAACTGCCAAGGG 4684
 QY 3967 gcaagccaagcagcgtcagcaagctatgtggatgtgagtgagtggaacaaagctctat 4026
 Db 4685 GCAAGCCAAGCAACAACTCAGCAGGCTATGTGGATGTGATGGCAATTAAGTCACTAT 4744
 QY 4027 gacagtacgataagaagtactatcaagtcaatgacaaggtcaagtgagacaaaacaaa 4086
 Db 4745 GACAGTACCGATAACAGTACTATCAAGCCAAATATGATGGCACAGTTGATAAACCAAA 4804
 QY 4087 gaagtggcaagacaaactgtgcgccagcccaaccccgatggcacattggctcaa 4146
 Db 4805 GAAGTTGCCAAGACAACTGTGTGCCAAGCCCAAAACCCAGATGGCATTTGGCTCAA 4864
 QY 4147 atgaatgcaaatcagtcatttaacaaagcaagtaaatgatgccaataaaagcaaggc 4206
 Db 4865 ATGAATGTCAAAATCAGTCAATTACAAGAACAAAGTAAATGATGCCAATAAAGCAAGGC 4924
 QY 4207 atcaatgaagacacgcctttatcaaaaggccttgaaagccgcgcgaagacacccaaacc 4266
 Db 4925 ATCAATGAAGACAAACGCCCTTTGTTAAAGGACTTGAAAAAGCCGCTTCTGATAACAAACC 4984
 QY 4267 aaaaacgcgcagtaactgtgggtgatttaaatgcccgttgcccaaacacgcgtgaccttt 4326
 Db 4985 AAAAAAGCCGACGTAACGTGTGGTGTATTAATTTAAATGCGCTTGCCCAACACCGCTGACCTTT 5044
 QY 4327 gcagggtacagggcacaacgcgtctaaactggcgagactttgaccatcaaaagtggg 4386
 Db 5045 GCAGGGATACAGCAACACCGGTAAAAACTGGGGAGACTTTGACCATCAAAAGTGGG 5104
 QY 4387 caaacagacaccaataagctaaaccgataataacatcggtgtgtgtagcaggtactgtgac 4446
 Db 5105 CAAACAGACACCAATAAGCTAACCGATAATAACATCGGTGTGGTAGCAGGTACTGTATGGC 5164
 QY 4447 ttaactgcaaaacttgccaaagcctaaacactcttaacagcgttaatgcaggtggcacc 4506
 Db 5165 TTCACTGTCAAACTTGCCAAAGACCTAACCAATCTTAACAGCCGTTAATGCAAGTGGCACC 5224
 QY 4507 agaattgatgaaagcactctctttttagacgcgaacgggtcaagcccaagcaaacacc 4566
 Db 5225 AAATTTGATGACAAAGCGGTCTTTTGTAGACTCAAGCGGTCAAGCCAAAGCAACACC 5284
 QY 4567 cctgtgtaagtgcacatgggctgggaactgggtggcaaacgcacatcagtaacalcggtgca 4626
 Db 5285 CCTGTGTAAGTGCCAAATGGCTGGACCTGGGTGGC----- 5320

QY 4627 gctgtgtgatgaacgatgcggtgaactttaagcagtttaagtggtgccaacacggtc 4686
 Db 5321 ----- 5320
 QY 4687 acaacctaacaacccaagaataactcaggtgcgtcattacccttctgtgtaacgcgtgc 4746
 Db 5321 ----- 5320
 QY 4747 aatggcaagcccatcaatggcacgcgatggcaagcccccaaaagccatcaagggcgccgat 4806
 Db 5321 ----- 5320
 QY 4807 ggtaaatactatcatcgcaacgcgcaacgcggtacgtggacaagaatggcaagcccatc 4866
 Db 5321 ----- 5320
 QY 4867 accgatgcggacaacacttgcgaactctggcagctcatggcaaaccccttgcaggtcat 4926
 Db 5321 ----- 5320
 QY 4927 caagtggcgaagccttaggcgcaactcagatgcacatcacccctaaaccaacatcaagtc 4986
 Db 5321 ----- 5320
 QY 4987 acttggccacaattgacacacccaaacacaggttaatgccaatgcaggcgcaagcccaagtc 5046
 Db 5321 ----- 5320
 QY 5047 ctgccagcctatcagcagcacagcaaaagtaagtgcgcagtgcaaaagatgtgtcaaat 5106
 Db 5321 ----- 5320
 QY 5107 gtaggctttaactgcagacacaaatcaacaatcaagtggacttgcgaagccctatgatacc 5166
 Db 5321 ----- 5320
 QY 5167 gtcaactttgtcaatggttacaggtgcgacatcaccaagcgtgcgtgagtggtgcacg 5226
 Db 5321 ----- 5320
 QY 5227 atgagtaacataccgctcaacacgcgccttagcgcagccgatgatgtggcaatgtgctt 5286
 Db 5321 ----- 5320
 QY 5287 atcaaaagccaaagatggttaagtcttacaagaagcagacacctcatgcacaaagcgtcacta 5346
 Db 5321 ----- 5320
 QY 5347 aaagcaggcaaatcagccagtgatgccaacactccaactggcttaagccttgttaacccc 5406
 Db 5321 ----- 5320
 QY 5407 aatgctgttaaggcagtcacaggcgatgcagtggtcttataataacttatcaaaaagcgta 5466
 Db 5321 ----- 5320
 QY 5467 ttttaaatccaagatggtacaactactaccacagtaagctctgtgagcagtatccaa 5526
 Db 5321 ----- 5320
 QY 5527 ggcaaaagataacagcagcatcacccctaaagatgggtgaaatgtgagcggttaagtc 5586
 Db 5321 -----AAGGTC 5326
 QY 5587 atcagcaatgtgggtaaaggaacaaagacacgcgcgtgccaatgtacaacagttaac 5646
 Db 5327 ATCAGTAATGTGGGCAAGGCAAAAGATACCGACGCTGCCAATGTACACACAGTTAAAC 5386
 QY 5647 gaagtacgcaactgttgggtcttctgtaaatgctggtaataacgcgtgacgcgcaatcag 5706
 Db 5387 GAAGTACGCAACTGTTGGGTCTTGTTGATGCTGGTAAATGATAACGC"TCACGGCAATCAG 5446

Qy	5707	gtaacattgcgcacatcaaaaaagaccocaaattcaggtlctcatctatacccgcaactgtc	5766
Db	5447	GTAAACAATTGCCACACATCAAAAAGACCCAAATTTCAGGTTTCATCATCTTAACCGCACTGTC	5506
Qy	5767	atcaagcagcgacgtactctgcggttaaggttaataacagatccgaaaaacttgcact	5826
Db	5507	ATCAAGCAGCGACGCTACTTGGCGGTAAAGGTAATAACGATACCGAAAACTTGGCCACT	5566
Qy	5827	gggtggttacaagtgggcgtggtataaagacgcgaacgctaacgcgatttaagcaaatgttt	5886
Db	5567	GGTGGTATACAAGTGGCGCTGCTATAAAGAGGCGCAAGCTAACGGCGATTTAAGCAATGTT	5626
Qy	5987	tgggtcaaaaccccaaaaagatggcgcaaaaaagcccttgcctgcgcacttataacgcgcga	5946
Db	5627	TGGGTCAAAACCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCACCTTATAACCGCGCA	5686
Qy	5947	ggtcagacaaactatttgccaaaccccgccagagaaccttgacagataaaatgaacaa	6006
Db	5987	GGTCAGACCAACTTTTGTACCAACAACCCCGCGAAGCCATTGACAGTAATATGAACAA	5746
Qy	6007	ggatccgctcttccatgtccaacgatggcaatcaagagcctgtggtacaaaggcgttaac	6066
Db	5747	GGTATCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGGTAAAGGCGGTAAAC	5806
Qy	6067	ggcattgactcaagtgcctcaggccaagcactcagtgggcga taggtttccaggccaaaggca	6126
Db	5807	GGCAATTGACTCAAGTGCCTCAGGCAAGCACTCAGTGGCGGATAGGTTTCCAGGCCAAGGCA	5866
Qy	6127	gatggtgaagccgcttgcctataggcagacaaacccaaagcaggaaccaactccatgcgc	6186
Db	5867	GATGGTGAAGCGCCCTGTGCCATAGCGAGCAAAACCCAAAGCAGGCAACCAATCCATCGCC	5926
Qy	6187	atcgggtgataacgcacaagccagcgcgatccaatccatcgcgcatcggtacagggcaatgtg	6246
Db	5927	ATCGGTGATAACCCACAAGCCAGCGGCGATCAATCCATCGCCATCGGTACAGGCAATGTG	5986
Qy	6247	gtaacaggtaagcactctggtgccatcgcgccaaagcactgttaagggttgataaacagt	6306
Db	5987	GTAGCAGGTAAAGCACTCTGTGTCCATCGCGCACCCCAAGCACTGTTTAAGCGTGATAACAGT	6046
Qy	6307	tacagtggtgtaataacaaccagttatcgaatgccactcagacccgatgctcttgggtg	6366
Db	6047	TACAGTGTGGGTAAATACAAACAGATTACCGATGGCATCAAAACCGATGTCTTTGGTGTG	6106
Qy	6367	ggcaataacatcacctgagccgaagtaactcggttgctttaggttcaaacctctgcgcatc	6426
Db	6107	GGCAATAACATCACCTGTACCGCAAGTAACTTCGGTTGCCTTAGGTTCAAACTCTGCCATC	6166
Qy	6427	agtcagggcacacgcagcgacacaaagccaaaaatctgcgcgcagcagcaggtacaacc	6486
Db	6167	AGTCAGGCGCACACGCGAGCACACAAGCCAAAAAATCTGACGCGCACAGCAGGTACAACC	6226
Qy	6487	accaacaggtgcaacaggtacggttaaaggcttgcctggcgaacaaacgggttggtgcg	6546
Db	6227	ACCACACAGGTGCACACCGGTACGGTTAAAGGCTTTGTGTGCAAAACGCGGGTTGGTGGC	6286
Qy	6547	gtctccgtgggtgcctcaggtgctgaagcgcgtatccaaaaatgggcagcaggtgaggtc	6606
Db	6287	GTCTCCCTGGGTGCCCTCAGTGCTGAACCCCGTATCCAAAATGTGGCAGCAGGTGAGGTC	6346
Qy	6607	agtgccaccagcaccgatcggttcaatggtagccagttgttacaagccaccccaaggcatt	6666
Db	6347	AGTGCCACCAAGCACCGATGCGGTCAATGTTAGCCAGTTGTACAAGGCCACCCAAAGCATT	6406
Qy	6667	gcaacgcgaaccaaataagcttgaccaatcgtatccaccacaaacgaaaaataaaggccaaatgca	6726
Db	6407	GCCAACGCAACCAATGAGCTTGACCATCGTATTCCACCAAAAACGAAAAATAAGGCCAATGCA	6466
Qy	6727	ggatattcatcagcgatggcgatggcggtccatgcccacaagcctacattcctggcgatcc	6786
Db	6467	GGGATTTTCACGCGATGGCGATGGGTGCCATGGCCCAAAAGCCTACATTTCTTGGCAGATCC	6526
Qy	6787	atggttaccgggggtattgcccacccacacagctcaaggtgcgggtggcgatggagactcgc	6846

Db	6527	ATGCTTACCGGGGTATATGCGACCCACAAACGGTCAAGGTGCGGTGGCAGTGGGACTGTGCG	6586
Qy	6847	aagctgcgataatgtcaatgggtatttataaaatacaatgggttcagccgataaccagaagc	6906
Db	6587	ANGCTGCGGATATATGCTAATGGGTATTTAAATCAATGTTTCAGCCGATACCCAAGGC	6646
Qy	6907	catgagggcgccagttggtgcaggttttctacttt	6942
Db	6647	CATGTAGGGCGGCAGTTGCTGCAGTTTTCTACTTT	6682
RESULT	6		
LOCUS	AR040716	6973 bp	DNA
DEFINITION	Sequence 1 from patent US 5808024.		PAT
ACCESSION	AR040716		
VERSION	AR040716.1	GI:5960079	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 6973)		
AUTHORS	Sasaki, K., Harkness, R. E., Loomore, S. M. and Klein, M. H.		
TITLE	Nucleic acids encoding high molecular weight major outer membra		
JOURNAL	protein of moraxella		
FEATURES	Patent: US 5808024-A 1 15-SEP-1998;		
source	Location/Qualifiers		
	1..6973		
BASE COUNT	2265 a 1555 c 1533 g	1620 t	
ORIGIN			
Query Match	37.2%;	Score 2580.6;	DB 9; Length 6973;
Best Local Similarity	66.9%;	Pred. No. 0;	
Matches 4803;	Conservative	0; Mismatches 1104;	Indels 1270; Gaps
Qy	1	atgaatcacatctataaagtcattctttaacaagccacagcacatcttatgtgcccgtggca	60
Db	542	ATGAATCACATCTATAAGTCATCTTTAAACAAGCCACAGGCACATTTATGGCAGTGCCA	601
Qy	61	gagtcgcgcaaatccccacagc-ggaaggagtagcagtagtaccgcaggaagcaggtggcgag	119
Db	602	GAGTACGCCAAATCCACACAGCGGGGGGGTAGCTGTCTACAGGCCAAGTTGGCAG	661
Qy	120	ctctctgtcatcgcgcgtgactggtgtgccacgctcgctatcctagtgtatgcgtgcagac	179
Db	662	TGTATGCACCTGTAGCTTTGCCCGTATATGCGGCGCTGCTGTCTCGTATCGGTGCAAC	721
Qy	180	gctcaatggcagtgcttatgctcaaa-----ataatagcaagatcgatttgct-----	228
Db	722	GCTCAGTGGCAGTGCTTTATGCTCAAAAAGATACCAACAATTCGCAANTTGGTGAACA	781
Qy	229	----accacagggcaacaatgacaatgcctcggtagcaatgaagcatccattgctattgg	284
Db	782	AAACAGCCAGACGCTCAGGCACCTGCCAAGGGCGGACGGTGTATCGAGCCATTTGCTATTGG	841
Qy	285	tagttctgtaaggcagcatgccaatcaagctattgctatcggtgtgtagcaaacaccagatcc	344
Db	842	TGAAATGCTAACGCACAGGGCGGTCAAGCATCGCATCGGTAGTAGTAA---TAAAC	898
Qy	345	tcgtaatacagcggtcaatcagaagcagcaggttcccacgcaaaaggtataagatccatgc	404
Db	899	TGTCATATGGGAAGCAGTTTGGATAGATAGGTACCGATGCTACGGGTCAAGATCCATCGC	958
Qy	405	catcgggtgtagtactgctgagggtagtgcctcgattgcccattggtagtgatgactt	464
Db	959	CATCGGTGGTGATGTAAAGGCTAGTGGTGATGCTCGATTGCCATCGGTAGTAGTACTT	1018
Qy	465	atatttgatagggaatagcactaacttaataatccaaatgg---tctcttagcactct	521
Db	1019	ACATTTGCTTCATCAGCATGTTATCTTCAACATCCGAAAGGTACTCTGATTAAAGTCT	1078

QY 522 tattcaaaaccatacagattattacgccaataatcagagactcaaatggtttctca---gaataa 578
 Db 1079 TATTAAGGCCCATTCAGTATTAAGAAATACGAAGCTCAAGAGTAATGATGTAATAA 1138
 QY 579 tagacacacagacagagaacagccagttactcagtgagaccatggcatatgcaaa 638
 Db 1139 TAGACGCAACACCGGAAGCGGACAGCCAGTACTGCAGTGGAGCCATGTCATATGCACA 1198
 QY 639 gggtaatttgcacacgcctttggtacacggttcaacagctgaaggcaacttctctggc 698
 Db 1199 GGGTCATTTTCACAGCCCTTTGGTACACGGGCAACAGCTAAAGTGCCTATTCCTTGGC 1258
 QY 699 agtagctcttacgcacgaagccgaagagagataatacatcgtctattggttcttaatagcaca 758
 Db 1259 AGTGGGTCTTGGCGCCACACAGCGAGGGCCAAATCTACAATCGCTATTTGGTTCATGATGAAC 1318
 QY 759 agctatcaattatggagactacgccttggtagacatactcagattgattggattaagg 818
 Db 1319 ATCTAGCTCGTTGGAGCGCATAGCCCTTGGTCGAGGTACTCGTCAAGTACAGGGCAG 1378
 QY 819 tattgccttaggttattggttctcagatccttataataataataataataataataataata 875
 Db 1379 TATTGCCCTAGGTCAAGGTTCTTGTCTACTCAGAGTGTATATAATCTTAGACCGGCCCTA 1438
 QY 876 agcctatgacagaagglaattgggttcaaacataaaatcgtctaaagccaccgggcaattg 935
 Db 1439 TACACCAATAACCCAGCCTAGACCCCAAGTTTCAAGCCACCAATAATACGAAGCGGG 1498
 QY 936 ttattttccattggttagtagcactatcaaggttaaaatcaatcgtcgggtgcagggtta 995
 Db 1499 TCCACTTTCCATTTGGTAGTAACCTCTATAAACGTAATAATCATCAATGTGCGTCAAGGTG 1558
 QY 996 tgaggataacagatcggttcaattgacacagctaaagcgtggagaaatctggcta---a 1052
 Db 1559 TAATAAACCGATCGGTCATATGTGGCACAGCTAGAAAGCGGTGGTGAAGTGGCTTAAGGA 1618
 QY 1053 qcgtcaaatlacttttaagggtagatataaacggtactggtttaaagaaactggggcga 1112
 Db 1619 GCCTAGAAATTTACTTTTCAGGGTGATGATAACACTACTGACGTAAAAATAGTTTGGATAA 1678
 QY 1113 gactttaaccattaaaggttggttagaccacgaagcggagcaagctaacgataataataacat 1172
 Db 1679 TACTTTTAACCTATTAAGAGTGGTG---CAGAGCCACAGCATTAACCGTAATTAATATATCGG 1735
 QY 1173 tgggtggttaacagataataactggtctgaaagtttaaacttgcataaaacccaagcgg 1232
 Db 1736 TGTGTTAAAGAGCTGATAATAGTGTCTGAAAGTTAAACTTGTCTAAACTTTAAACAA 1795
 QY 1233 tcttgaacacagttagcaccacaaaccttaaccgcccagcggagaaagttacgtag---gtag 1289
 Db 1796 TCTTACTGAGGTGAATCAACTACATTAATTAATGGCCCAACCCACAGTTAAGGTAGTAGTAG 1855
 QY 1290 tggtaataaacaccgctgagctacaaaagcgggtggttttaacctttacc---caacaac 1343
 Db 1856 TAGTAGTACTACAGCTGATTAATGAGTAGTAGTTTAACCTTTACCCAGCCCAATACAGG 1915
 QY 1344 aaatgaacacagacacaaacccgtctatggtcactgtagggttaagtttactgataatc 1403
 Db 1916 CAGTCAAAAGCACAGCAAAACCGTATGCGGTTAATGGGGTGAAGTTTACTAATAATGC 1975
 QY 1404 taatagccactgaa---gatactactcgtatcaccaagaataaaattggttttagcaa 1460
 Db 1976 ACAAAACACAGCAATTCGGCCTACTCTGCTATTTACAGAGATAAATTTGGCTTTGTCTCG 2035
 QY 1461 taaagctggttacagttgatgaaaaaaccttattcttgataaagacaaagctaaag--- 1516
 Db 2036 AGATGGTGTATGTTGATGAAAAACAAGCACCATTATTTGGATAAAAAACAACCTTAAGTGGG 2095
 QY 1517 -----t 1517
 Db 2096 TAGTGTTCGAATTAACATAGACATGGCATTTGATGTCAGGTAAATAAAGATCAGTAATCT 2155

QY 1518 tgcaacagcaccttaaaacacggttggtgactgttaataaacacacattgggtggtagcaa 1577
 Db 2156 TGCCAAAGGTAGCAGTGTCTAACGATGCGGTTACCATCGAACAGCTCAAAAGCCGCCAAGCC 2215
 QY 1578 taacaaatccaagtcggtgctgctgattggtcattaaattggccgagtgtaattgtaattatc 1637
 Db 2216 TACTTTAAAGCGACGGCTGGCATCATGTCACACCTACTGAAATATTCAGTTGATGCTTAA 2275
 QY 1638 aaatgcgcgaataattcgcactactcgtattaccgaagaggaattgggttctgctgacgc 1697
 Db 2276 GAGTGGCAATGTTTACCCGCCCACTTACAACTATGGCGTCAAAACCCAGCTTAACAG 2335
 QY 1698 tgatggtaaaagttgataa----- 1715
 Db 2336 TGATGGCACATAGTCAATAATTTAGTGTAAAGGTAGTGGTACGAACAATAGCTTTAGTTAC 2395
 QY 1716 -----aaagtcaccatatttggataaaaaaacaacttcaagttgggtgggtgt 1760
 Db 2396 CGCGAACAATTTGGCAAGCTATCTTAATGAAGTCATCTGACGCTGACAGTGTCTCTACA 2455
 QY 1761 taaattaccaaagacagtggtcatttaagcaggtgatacaaaagatcagtaattgtaaaaga 1820
 Db 2456 AAGCTTTTACCGTTTAAAGAAAGACGATGATGACGCCAACGCTATCACCGTGGCTAAAGA 2515
 QY 1821 tgcaacggagacataccgagtcagtcacttataaacagcttataaacagtcgaacaaagcgc 1880
 Db 2516 TACGACAAAAAATGCCGCGCAGTCAGCATCTTAAACTCAAGGTAAACACGGTCTTAAC 2575
 QY 1881 cgaaggtgcctacaa-----agcttctctattcgtgatgaaaaaggtcagga 1928
 Db 2576 GGTGCTACCAAAAAGATGTACGGTTACCTTTGGGCTTAGCCNAGATAGCGTCTGAC 2635
 QY 1929 attacgattgtaacttcttatttcttaattgtaataacacccaataacccttgagaccatcac 1988
 Db 2636 CATTTGGCAAAAACCCCTTAAACACAGATGGGTGACTTTAAAGATACCAACGACAAAT 2695
 QY 1989 ctttgcaggtgaaacgcatcagtatcagcaatgacatagcaccaggttaaaagtcgaag- 2047
 Db 2696 CCAAGTCGGTCTAATGGCATTAATTTACTTAATGTGAATGTTAGTATCCAGGTACTGG 2755
 QY 2048 ---ttggtattgaccacaaatcaatggtctcacacgcgtcgaagctgacctgggttagcgataa 2105
 Db 2756 CATTCGAATACCGCTCGCATTTACCAGAGATAAATTTGGCTTTGCTGTTCTGTATGGTGC 2815
 QY 2106 agatggttaaaactcaattggttattgagcaag----- 2137
 Db 2816 AGTTGATACAAACAACCTTATCTTTGATACAGCAAGCTACAAGTTGGCAATGTTAAGAT 2875
 QY 2138 -----tggttagcgttaacgcaccccaaaaaacatcattagagagattgtcccaaacactgcc 2192
 Db 2876 TACCAACACTGGCATTAACGCAAGGTGGTAAAGCCCATCACAGGGCTGTCCCAACACTGCC 2935
 QY 2193 tagcattaccaatcaggtggtgctagcaccacagaaagggcgaatacaataccacagcga 2252
 Db 2936 TAGCATTCGGCATCAAGTAGC---CGCAACATAGACTGGGCAATACAAATCCA---AGA 2989
 QY 2253 cgaagacaaatccaaagccgcaggtatcggtgatatataataatacaggtcttaacctaaa 2312
 Db 2990 CAAAGACAAATCCAACGCTGCCAGCATTAATGATATATTAATACAGGCTTTAAACCTAAA 3049
 QY 2313 aaataatagcaactcgttgggttcttctccacttataaactgttgcacttctatcgatgg 2372
 Db 3050 AAATAATAACAACCCCATTTGACTTTCTCCACTTTATGACATTTGTGACATTTGCCAATGG 3109
 QY 2373 caatgcacacacccgttaagtaacttacgtgataaatacaaaccaatacaaacagtaactta 2432
 Db 3110 CAATGCCACCCGCCACAGTAACCCCATGATACCGCTAACAAACCCAGTAAAGTGTGTATA 3169
 QY 2433 tgatgtcaattgtgtagagaaaaacattgaaactcacagcgataatggtgagacaacaa 2492
 Db 3170 TGATGTGAATGTGGATGATACAAACCATTCATCTAACAGGCACCTGATGACAA---TAAAAA 3226
 QY 2493 aattggcgtcaaaacaccacactgaccacaacaaatgctaaatgggttaaagcaacc----- 2547

QY 4686 caacacccataaacaacaaagtaactcaggtgctcattaccccttctggttaacccgatgc 4745
Db 5322 -----
QY 4746 caatggcaagcccatcaatggcaccgatggcaagcccaaaaaagccatcaagggcgccga 4805
Db 5322 -----
QY 4806 tggtaaatatactatcacgcaacgcaacggtcgtaacctgtggcaaaagatggcaagcccat 4865
Db 5322 -----
QY 4866 caccgatgggacaaaacttgccaattctggcagctcatggcaaaccccttgatgcaggTca 4925
Db 5322 -----
QY 4926 tcaagtgggcaagcctaggcggaactcagatgccatcccttaacccaatcaagtc 4985
Db 5322 -----
QY 4986 cacttggccaaaatlgacacacaaaacacacaggttaatgccaatgcaggcaagcccaag 5045
Db 5322 -----
QY 5046 tctgccagcctatcagcagcacagcaagtaatgctgcagtgctcaagatgtgtgtataa 5105
Db 5322 -----
QY 5106 tgtaggctttaacttgagaccacatacaatacgaagtggactttgtcaagccctatgatatac 5165
Db 5322 -----
QY 5166 cgtcaacttltgtaattgttacaggtgccgacatacaagcgtgcgtagTgctgatggcac 5225
Db 5322 -----
QY 5226 gatgataacatcacctgcacacccgcttagcagcagcgtatgatggcaatgtct 5285
Db 5322 -----
QY 5286 tatcaagccaaagatgtaagtctctacaagcagacgacacctcatgccaaaagcgtcact 5345
Db 5322 -----
QY 5346 aaaagcaggcaaatcagccagtgatgccaaaactcccaactggtcttaagccttlttaaccc 5405
Db 5322 -----
QY 5406 caatgctggttaaaagcagtacagcgtatgcagtggtctcttaataaactatcaaaaagcgt 5465
Db 5322 -----
QY 5466 atttaaatccaaagatggtacaactactactaccacagtaagctctgatggcatatcca 5525
Db 5322 -----
QY 5526 aggcgaagataacagcagcatcacccctaaagcaagatggcgtgaatgtaggcgttaaggt 5585
Db 5322 -----
QY 5586 cactagcaatgtgggttaaggcacaaaagacacgcagcgtgccaatgtatacaacagtttaa 5645
Db 5327 CATCAGTATGTGGGCAAGGCACAAAAGATACCGACGTGCCAATGTACAAACAGTTAAA 5386
QY 5646 cgaagtacgcaacttgggtcttgggttaattgctggtaataataacgctgcagcgaatca 5705
Db 5387 CGAAGTACGCAACTTGTGGGTCTTGGTAACTGCTGTTAATGATAACGCTGACGGCAATCA 5446
QY 5706 ggttaacattgccgacatacaaaaaagacccaaattcaggttcatcatcctaaccgactgt 5765
Db 5447 GGTAAACATTGCCGACATCAAAAAAGACCCAAATTCAGTTTCATCTCACTAACCCGACTGT 5506

QY 5766 catcaaaagcaggcacggtactctggcggttaaggtataacgataccagcaaaaacttgccac 5825
Db 5507 CATCAAAAGCAGGCACGGTACTTGGCGGTAAAGTAATAACGATACCGAAAACTTCCAC 5566
QY 5826 tgggtgtgataagtggtgataaagacgcaacgctaaacgcttaacgcttaagcaatgt 5885
Db 5567 TGGTGGTATACAAGTGGCGGTGGATAAAGACGGCACGCTTAACGGCGANTTTAAGCAATGT 5626
QY 5886 ttgggtcaaaaccccaaaagatggcagcaaaaaagccctgctgcacctataaacccgc 5945
Db 5627 TTGGGTCAAAACCCANAAAGATGCGACGAAAAAGCCCTGCTGCCACTTATTAACGCCG 5686
QY 5946 aggtcagaccaaactatttgaccacaaccccgagagaagccattgacagaataaataaca 6005
Db 5687 AGGTCAAGACCAACTATTGTACCAACAACCCCGCAAGACCATTTGACAGATAAATAAGCA 5746
QY 6006 aggtatccgctcttccatgtcaacgatggcaatacaagagcctgtgtaacaaagggcgtaa 6065
Db 5747 AGGTATCCGCTTCTTCATGTCAACGATGCAATCAAGAGCCTGTGCTCAAGAGGCGTAA 5806
QY 6066 cggcattgactcaagtgcctcaggcaagcactcaagtggcgatagggttccaggcccaaggc 6125
Db 5807 CGGCATTGACTCAAGTGCTCAGGCAAGCACTCAAGTGGCGATAGGTTTCAGGCGCAAGGC 5866
QY 6126 agatgtaagcccgcttgccatagcagacaaaacccaaagcagcaacccaatccatcgc 6185
Db 5867 AGATGGTGAAGCCGCCGTTGCCATAGCGACAAACCCAAAGCAGCAACCAATCCATCGC 5926
QY 6186 catcgtgtataacacacacgagcggcgatcaatcaatcgatccatcggtacagcgaatgt 6245
Db 5927 CATCGGTGATAACGCACAAAGCGCGGCGATCAATCCATCGCCATCGGTACAGGCAATGT 5986
QY 6246 ggttaacagtaagcaactctggtccatcgcgacccagcactgttaaggtgataacag 6305
Db 5987 GGTACAGAGTAAGCACTCTGTTGCCATCGGCAACCCAAAGCACTGTTAAGGCTGTATACAG 6046
QY 6306 ttacagtggtggttaataacacacagctttatcgatgccactcagaccgctgtctttgggt 6365
Db 6047 TTACAGTGTGGGTATTAACACCACTTTACCGATGCCACTCAAAACCGATGCTTTGGTGT 6106
QY 6366 gggcaataacatacccgtagccgaaagttaactcggtcccttaggttcaaaactctgccc 6425
Db 6107 GGGCAATAACATACACCGTGACCGAAAGTAACTCGGTTGCCCTTAGGTTCAAACTGCGCAT 6166
QY 6426 cagtcagcacacacgagcagcacacaaagcccaaaaaactcagcgcacagcaggTacaac 6485
Db 6167 CAGTCAGGCACACAGCGACGACACAAAGCAAAAAATCTGACGGCACAGCAGGTACAAAC 6226
QY 6486 caccacagcaggtgcaacaggtacggttaaggtttgctggacaaacggttgggtgc 6545
Db 6227 CACCACAGCAGGTGCNAACCGGTACCGTTTAAAGGCTTTGCTGGACAAACGGCGTTGGTGC 6286
QY 6546 ggtctcgtgggtgctcaggtgctgaacggttaLccaaaaatgtggcagcaggtgaggt 6605
Db 6287 GGTCTCCGTGGGTGCTCAGGTGCTGAACCGCGTATCCAAAAATGTGGCAGCAGGTGAGGT 6346
QY 6606 cagtcacacagcagcagatgctggtgtaagcaggtgtgatacaaaagccaccccaaggtat 6665
Db 6347 CAGTGCACACAGCAGCAGTGCAGTCAATGGTAGCCAGTTGTACAAAGCCACCAAGCAT 6406
QY 6666 tggcaacgcaaccaaataagcttgacctatccaccacaaacgaaaaataaagcccaatgc 6725
Db 6407 TGCCAACGCAACCAATAGAGCTTGACCATGTTATCCACAAANACGAANAATAAGGCCANTGC 6466
QY 6726 agggatttcacatcagcagatggcggtccatggcacaagcctacattccttggcagatcc 6785
Db 6467 AGGGATTTCATCAGCATGGCGATGGCGTCCATGGCCACAAGCCTACATTCCTTGGCAGATC 6526
QY 6786 catgggttacccgggggtatttgcaccccaacacggTcaaggtgctggtgagcagctgc 6845
Db 6527 CATGGTTACCGGGGTATTGCCACCCACACAGCGTCAAGGTGCGGTGGCAGTGGGACTGTC 6586
QY 6846 gaagctgctcggaataatgggttcaatgggtatttaaaatacaatgggttcagcgcgatacccaagg 6905


```

Db 6587 GAAGCTGTGGGATAATGGTCAATGGGTATTTAAATCAATGGTTACGCCGATACCCAAAG 6646
QY 6906 ccaatgtagggcgaggttggtgaggttttcaacttt 6942
Db 6647 CCATGTAGGGCGGCAGTTGGTGCAGGTTTCACTTT 6683

RESULT 7
AX079913 LOCUS Ax079913 6973 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 1 from Patent WO0107619.
ACCESSION AX079913
VERSION AX079913.1 GI:13159436
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 6973)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of
moraxella
JOURNAL Patent: WO 0107619-A 1 01-FEB-2001;
CONAUGHT LABORATORIES LIMITED (CA)
FEATURES
    source
        1..6973
        /organism="Moraxella catarrhalis"
        /db_xref="taxon:480"
BASE COUNT 2265 a 1555 c 1533 g 1620 t
ORIGIN

Query Match 37.2%; Score 2580.6; DB 10; Length 6973;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 4803; Conservative 0; Mismatches 1104; Indels 1270; Gaps 29;

QY 1 atgaatcacatctataaagtcatcttttaacaaagccacagggcacattatggcgtggca 60
Db 542 ATGAATCACATCTATAAAGTCATCTTTTAAACAAAGCCACAGGCACATTTATGGCAGTGGCA 601
QY 61 gagtgcacaaatccacagc-ggagggagtagcagtagtaccgcagagacagatggcag 119
Db 602 GAGTAGCCAAATCCACAGCACGCGGGGGGGTACTGTGTACAGGCGCAAGTTGGCAG 661
QY 120 ctctctgtctacgcgctgactgtgttgcacgctcgtctatctctctagtgtgac 179
Db 662 TGTATGCACTCTGAGCTTTGCCGGTATTGCCGCGCTCGCTCTCTGTGATCGTGCAC 721
QY 180 gctcaatggcagtgcttatgctcaaa-----ataatagcaagatcgcatttggt---- 228
Db 722 GCTCAGTGGCAGTGGCTATGCTCAAAAAAAGATACCAACATATCGCAATTGGTGAACA 781
QY 229 ----accacagcacaatgacatgacctcgctagcaatgaagcagcatctctattgg 284
Db 782 AACCCAGCCAAAGCTCAGGCACTGCCAGCGGAGGCTGATCGAGCCATTCATTTGG 841
QY 285 tagcttctgtaagggcacatgccaatcaagctattgtatctgtgttgtagcaaacagatcc 344
Db 842 TGAATAATGCTAACGCACAGGGCGGTCAAGCCATCGCCATCGGTAGTAGTAA---TAAAC 898
QY 345 tctaatcaagcggcttaatacagaagcagggttcccacgcccacaaaggttaagagtcctatcgc 404
Db 899 TGTCAATGGAAGCAGTTTGGATGAAGATAGGTACCCGATGTACGGGTCAAGAGTCCATCGC 958
QY 405 catcgttggtgtagtactgaggtgtagtgcctcagattgcatggttagtgaactt 464
Db 959 CATCGGTGGTGTGATGAAGGCTAGTGGTGAATGCTCGATGCTCGTGGTGGTGGTGGT 1018
QY 465 atattggataggaatagcactaactctaaataatccaaatgg---tctcttagcactct 521
Db 1019 ACATTGCTTGATCAGCATGGTAACTCTTAAACATCCGAAAGGTACTCTGTATTACGATCT 1078

```

```

QY 522 tattcaaaaccatacacagtattacgcccacaaatacacgagactcaaatgggtctca---gaaata 578
Db 1079 TATTACGGCCATGCGAGTATTTAAAGAAATACGAAGCTCAAGAGTAATCATGTATGATAATA 1138
QY 579 tagacacacagcagaagaagacacgccagctactgcagtgaggagccatgcatatgcaaa 638
Db 1139 TAGACGCACAAACCGCAAGCGGACACGCCAGTACTGCTAGTGGGAGCCCATGTCTATATGCACA 1198
QY 639 gggctattttgccaaagccttttggttacacggttcaacagctgaagggcaactatccttggc 698
Db 1199 GGTCATTTTCCACGCCCTTTGGTACACGGGCAACAGCTTAAAGTGCCTATTTCCTTGGC 1258
QY 699 agtaggtcttaccgcccacaaagcgaataatacaaatcgctatgggttcttaagtacaa 758
Db 1259 AGTGGGTCTTTGCCGCCACAGCCGAGGGCCAACTCTACAATCGCTATTGTTCTGTGCAAC 1318
QY 759 agctacaattatggagcactagccttggtcagatactcagctgattgattgattacgg 818
Db 1319 ATCTAGTCGTTGGAGCGGATAGCCCTTTGGTGCAGGTACTCGTCTCAGCTACAGGGCAG 1378
QY 819 tattgcctaggttattggttctcagatccttaataataataataataata---ataataa 875
Db 1379 TATTGCCCTAGGTCAAGGTCTCTGTCTACTCAGAGTGATAATAATTCTAGACCGGCTA 1438
QY 876 agcctatgtaccagaaggtaagtgggtcaaacataaaaatcgtctaaagccaccggcaatgg 935
Db 1439 TACACCAAAATACCCAGGCACCTAGACCCCAAGTTTCAAGCCACCAATAATAACGAAGCGGG 1498
QY 936 ttattttccattggttagtagcactatcaagcgttaaatcatcatcgtcgtcaggtta 995
Db 1499 TCCACTTTCCATTGGTAGTAACCTCTATCAACGCTAATAANTCATCAATGTCGGTGCAGGTGT 1558
QY 996 tggagataccgctgctcaatgtggcacagctaaaagcgttgagaaatctggcta---a 1052
Db 1559 TAATAAACCGATGCGGTCAATGTGGCACAGCTAGAAAGCGGTGCTGAAGTGGCTAAGGA 1618
QY 1053 gctcaaatctacttttaaggggtgatataacggtactgctgcttaagaataaaactgggagcga 1112
Db 1619 CGCTAGAAATTTACTTTTTCAGGGTGATGATAACAGTACTGACGTAAAAATAGTGTGTGATAA 1678
QY 1113 gactttaaccattaaagggtgtagaccccaagcgcaagcgaagcgaagcgaagcgaagcgaagcga 1172
Db 1679 TACTTTAACTATTAAAGGTGGTG---CAGAGACCAACGCGATTTAACCGATATATATATCGG 1735
QY 1173 tgggtgggtaacagataataactgctgctgaaagtttaaaacttgcataaaacccctaaagcgg 1232
Db 1736 TGTGGTAAAGAGCGCTGATAATAATAGTGGTCTGAAAGTTAAACTTGTCTAAACACTTTAAACAA 1795
QY 1233 tcttgaacaagtttagcaccacaaacccctaacccgacgagaaaggttacggtag---gtag 1289
Db 1796 TCTTACTGAGGTGAATACAACTACATTTAAATGGCCACACACAGTTTAAAGGTAGGTAGTAG 1855
QY 1290 tggtaataacacccgctgagctacaaagcgtgtttaacctttacc---caacaac 1343
Db 1856 TAGTAGTACTACAGCTGAATATTATTGAGTGATAGTTTAACTTTTACCCTTTTACCAGCCCAATACGG 1915
QY 1344 aaatgcaagcacagacacaaacccctctatggcactgtagggcttaagtttactgataattc 1403
Db 1916 CAGTCAAAAGCACAAGCAAAACCCGCTATATGCGTTAATGGGTGAAAGTTTACTAATAATGC 1975
QY 1404 taatacggcaacttga---gatactactctatcaccacaaagataaaatgggttttagcaa 1460
Db 1976 AGAAACAACAGCAGCAATCCGCCACTACTCTGATTTACCCAGAGATAAAATTTGGTTCGTCG 2035
QY 1461 taaagctgggtacagttgataaacaacaccccttatcttgataaagacgaagcgaagcgaagcga 1516
Db 2036 AGATGGTGATGTTGATGAAGAAACAAAGCCACCATATTGCGATTAATAAACAACTTAAAGTGGG 2095
QY 1517 -----t 1517
Db 2096 TAGTGTGCAATTACCATACACAATGCGCATTTGATGCGAGGTAATAAAAAAGATCAGTAATCT 2155
QY 1518 tggcaacagcacccctaaacacacacgggtggcttgactgtaataacacacccattgggtgtagcaa 1577

```


Qy	3609	cagcattggtgagtgtctctaaacgcagggctttaacttgcgaagccaattggtgaagcgggttga	3668
Db	4328	CAGCATTTGTTGATGTGCTTAAGCGCAGGCTTTAACTTGAAGCGCAATGGTGAAGCGGTGA	4387
Qy	3669	cttctgtccacttatgacacgttgtactttatcatgctgacgaatgccaccaccctctaagt	3728
Db	4388	CTTTGTCTCCACTTATGACACCGTCACTTTGCCGATGGCAATGCCACACCGCTAAGGT	4447
Qy	3729	gacctatgatgacacaagcaaacaccagtaaaagtggctctatgatctcaatgtgataataa	3788
Db	4448	GACCTATGATGACACAGCAAAACACAGTAAGTGTCTATGATGTCAAATGTGGATGATAC	4507
Qy	3789	aaccattgaagtgaacagtgtataaaaaacttggcgtcaaaaaccaccacacactgacccaaaac	3848
Db	4508	AACCATTTGAAGTTTAAA---GATAAAAAACTTTGGCGTAAAAACACCACTATGACCAAGTAC	4564
Qy	3849	aagtgtcaatggttaatgca---accaaaattagtcgcccgatggcgatgcccttgttaa	3905
Db	4565	TGGCACAGGTTGCTAATAAATTTTCCCTTAAGCANTCAACTACTGGCGATCGCTTGTCAA	4624
Qy	3906	agccagtgtatcgcaccacatctaataccttggctggcgacatccaaaccgcgcaaaagg	3965
Db	4625	GGCCAGTGATATCGTTGCTCATCTAAACACCTTATCTGGCGACATCCAACTGCCAAAGG	4684
Qy	3966	ggcaagccaaagcagcgtcagcaagctatgtgagtctgatggcaaacaaagtcactcta	4025
Db	4685	GGCAAGCCAAAGCAACAACCTCAGCAGGCTATGTGGATGCTGATGGCAATTAAGTCAATCTA	4744
Qy	4026	tgacagtaccgataaagaagtactatcaagtcaatgacaagggtcaagtgtagacaaaaacaa	4085
Db	4745	TGACAGTACCAGATAACAGTACTATCAAGCCAAAATAATGATGGCACAGTTGNTAANACCAA	4804
Qy	4086	agaagtgtccaaagacaaactggtgcgccaagcccaaaccccgatggcacattggctca	4145
Db	4805	AGAAAGTTGCCAAAGACAAACTGGTGCCTCCAGCCCAACCCAGATGGCACATTTGGTCA	4864
Qy	4146	aatgaatgtcaaatcagtcattaaacaagagcaagttaataatgccataaaagaacagg	4205
Db	4865	ATGAAATGTCAAAATCAGTCAATTAACAAGAACAGTAATGATGCCAATAAAGCAAGG	4924
Qy	4206	catcaatgaagacaacgcctttatcaaaaggccttgaaacgcgcccaagacacacaaac	4265
Db	4925	CATCAATGAAGACAACGCCTTGTGTTAAAGGACTTGAAAAGCGCGTCTTGATTAACAAAC	4984
Qy	4266	caaaaacgcgcgtaactctgtgggtgatttaaatgcgctgtcccaaacacgcgtgacctt	4325
Db	4985	CAAAAACCGCGCAGTAACCTGTGGGTGATTTAATGCCGCTTGCCCAAAACCGGTGACCTT	5044
Qy	4326	tgcaagggtatcaggcacaaacgcgtlaaaaaactggcgagacttggacatcaaaagttg	4385
Db	5045	TGCAGGGGATACAGGCACAACGGCTTAAAAAATCTGGCGAGACTTTGACCATCAAAAGTGG	5104
Qy	4386	gcaaacagacaccaataaagctaacccgataataaacatcggtgtgtgtagcaggtactgatg	4445
Db	5105	GCAAAACAGACCAACAAAGCTAACCGATAATAACATCGTGTGTGTAGCAGGTACTGATGG	5164
Qy	4446	cttcaactgtcaacttgcacaagacctaaccaatcttaacagcgttaatgcaggtggcac	4505
Db	5165	CTTCACCTGTCAAACTTGGCCAAAGACCTAACCAATCTTTAACAGCGTTAATGCAAGGTGCAC	5224
Qy	4506	cagaattgataaaaaggcatctcttttgtagacgaacaggtcaagccaaaagcaaacac	4565
Db	5225	CAAAATTTGATGACAAAGGGGTGCTTTTGTGTAGACTCAAGCGGTCAAGCCAAAGCAACAC	5284
Qy	4566	cccttgtctaaagtgcacaatggcctggacccgtggcgcaaacgcgcatcagtaacatccgtgc	4625
Db	5285	CCCTGTGCTTAAGTGCCAATGGGCTGSACCTGGGTGGC-----	5321
Qy	4626	agctgtgtgataaacyatgcggtgaacctttaagcagtttaatgaagttgccaacacggt	4685
Db	5322	-----	5321

Qy	4686	caacaaactaaacacaaagtaactcaggctgcgtcattaccctcttgctgtaaaccgatgc	4745
Db	5322	-----	5321
Qy	4746	caatggcaagcccatcaatggcaccgatggcaagcccaaaaaagccatcaaggggccga	4805
Db	5322	-----	5321
Qy	4806	tggtaaatactatcacgccaaacgcaacgcgctacctgtggacaagatggcaagcccat	4865
Db	5322	-----	5321
Qy	4866	caccgatgcggacaaacttgcgaatctgcgactcattgcgaaccccttgatgcagtgca	4925
Db	5322	-----	5321
Qy	4926	tcaagtgtggcaagcctagcgcggaactcagatgccatcacctcaaccaacatcaagtc	4985
Db	5322	-----	5321
Qy	4986	cacttggccaaattgacacacacacacacaggaatgcgaatgcagggcaagcccaag	5045
Db	5322	-----	5321
Qy	5046	tctgccagcctatcagcagcagcagcaaaagtaatgctgccagtgctcaaatgtgctaaa	5105
Db	5322	-----	5321
Qy	5106	tgtaggctttaacttcagaccgaatcacaaatcaagtggactttgtcaaaagcctatgatac	5165
Db	5322	-----	5321
Qy	5166	cgtaactttgtcaatggtacaggtgcgacatcacaaagcgtgcgtagtgctgtagggcac	5225
Db	5322	-----	5321
Qy	5226	gatgagtaacatcacocctcaacaccgccttagcagcagccgatgatggtgcaatgtgct	5285
Db	5322	-----	5321
Qy	5286	tatcaagccaaagatggtgaattctacaagcagacacctcatgccaaacggctcaact	5345
Db	5322	-----	5321
Qy	5346	aaagcaggcaaatcagccsgtgatgccaaaacctccaactggtctaaagcctgtttaaccc	5405
Db	5322	-----	5321
Qy	5406	caatgctggtaaggcagtacaggcgatcgatggcctcttaataacttatcaaaagcgtt	5465
Db	5322	-----	5321
Qy	5466	atttaaatccaaagatggtacaactactaccagtaagctctgatggcatcagtatcca	5525
Db	5322	-----	5321
Qy	5526	aggcaagataacagcagcatcaccttaagcaaaagatggctggaatgtaggcggttaagtt	5585
Db	5322	-----	5326
Qy	5586	catcagcaatgtgggtaaaggcacaagaacaccagcgtgcgaatgtacacaggttaaa	5645
Db	5327	CATCAGTAATCTGGGCAAGGCCAAAAAGATACCGACGCTGCCAATGTACAACAGTTAAA	5386
Qy	5646	cgaagtacgaaactttgtgggtcttggttaatgctggtgaatgatcaacgctgcagccaatca	5705
Db	5387	CGAAGTACGCAACTTGTGGTCTTGGTAAATGCTGGTAAATGATAACGCTGACGGCAATCA	5446
Qy	5706	ggtaacattggcgacatcaaaaaacccaaattcaggttcaatcatcttaacccgactgt	5765
Db	5447	GGTAACATTCGCCGACATCAAAAAGACCAAAATTCAGGTTTCATCATCTAACCCGACTGT	5506
Qy	5766	catcaagcagcagcagctactttgcggttaaaqgtaataacatcacccgaaacacttgcac	5825

[illegible]

Db	538	GCCTATTTCCTTGGCAGTGGGTCTTGGCGCCACAGCCGAGGCCAATCTACAAATCGCTATT	597
Qy	745	ggttctaatacacaagctatcaattatgagcaactagcccttggtcagagataactcgaagt	804
Db	598	GGTCTGATGCACATCTAGCTCGTTGGGAGGATAGCCCTTGGTGACAGGTACTCGTGCT	657
Qy	805	gatttgattacggttatgccttaggttatggtcttcagataccttaataataataa	864
Db	658	CAGCTACAGGGCAGTATTGCCCTAGTCAAGGTCTGTGTCTCACTCAGAGTGATAAAT	717
Qy	865	aaata--ataataaagcctatgaccagaagtaatgggtcaaacataaaatgctctaaa	921
Db	718	TCTAGACGGCCTATACACAAATCCAGGCACCTAGACCCCAAGTTTCAGGCCACCAAT	777
Qy	922	gccaccggcaatggtttattttccattggtagtagcactcaagcgtaaaaatcatcaat	981
Db	778	AATACGAAGCGGGTCCACTTCCATTGGTAGTAACTCTATCAACAGTAAATCATCAAT	837
Qy	982	gtcggtcaggttatgagataccgatcggttcaatgttggcacagctaaagcgggtgag	1041
Db	838	GTCGGTCCAGGTGTTAATAAAAAACCGATGCGGTCAATGTGGCACAGCTAGAAAGCGGTGGTG	897
Qy	1042	aatctggcta---agcgtcaaatattctttaaagggtgaigataacggtactgcggttaag	1098
Db	898	AAGTGGGCTAAGGAGCGGTAGAATTACTTTCAGGGTGATGATAACAGTACTGACGTAAAA	957
Qy	1099	aaaaaactggcgagactttaaccattaaagtggtgagaccgaagcggacaagctaac	1158
Db	958	ATAGGTTTGGATTAATACTTAACTATTAAAGGTGGTG--CAGAGACCCACGCAATTAACC	1014
Qy	1159	gataataatacaattgggttggtaacagataataactggcttgaagtttaaaacttgct	1218
Db	1015	GATAATAATATCGGTGTGTTAAAGAGGCTGATAATAGTGTCTGAAAGTTTAAACTTGTCT	1074
Qy	1219	aaaaacctaagcgggtcttgaacagtttagcaccaaaacaccttaaccgccagcgagaagtt	1278
Db	1075	AAAACTTTAAACAATCTTACTGAGGTGAATACCACTACATTAATATGCCACCAACACAGTT	1134
Qy	1279	acgttaggttagtgt---aataaacccgtgagctcacaaagcgggtttaaacccttacc	1335
Db	1135	ANGGTAGGTAGTAGTAGTACTACAGCTGAATTAATTGTAGTGATAGTTTAACTTTTACC	1194
Qy	1336	c-----caacaacaaatcgagcagacagaaacccgtctatggcactgatgggtctaa	1389
Db	1195	CAGGCCAATACAGGCAGTCAAAGCAAGCAAGCAAAACCGTCTATGGCGTTTAATGGGGTCAAG	1254
Qy	1390	tttactgataattctaatcagccacttgaa---gatactactctgatacacaagataaaa	1446
Db	1255	TTTACTAATAATGACGAAACACAGCAGCAATTCGGCACCTACTCGCTATTACCAGAGATAAA	1314
Qy	1447	atggtgtttagcaataaagcttgtagttgatgaaacaaaccttattcttgataaagac	1506
Db	1315	ATTGGCTTTGCTCGAGATGGTGATGTTGATGAANAACAGCACCACATATTGGATAAAAA	1374
Qy	1507	aagctaaagtgtg-----	1520
Db	1375	CAACTTAAAGTGGGTACTGTCCAATTACCATAGACAATGGCAATTGATGCAAGTAATAAA	1434
Qy	1521	-----caacagcaccttaaacacagtggtgtgacttgacttgtaataaacacc	1563
Db	1435	AAGATCAGTAATCTTGCCAAAGGTAGCAGTGCTAACGATGCGGTATCCATCGAACAGCTC	1494
Qy	1564	atggtgttgagataaacaacaaatccaagtgcggtgctgattggcattaaatttgcgcatgtg	1623
Db	1495	AAAGCCGCCAAGGCTTACTTTAAACGACGCGCTGGCATCGATGTCACACCTTACTTGAATA	1554
Qy	1624	aatgttaagtatcaaatgcgcgcaaaatcgggcactactcgttattaccgaagaggaatt	1683
Db	1555	TCAGTTGATGCTTAAGAGTGGCAATGTTACCGCCCCCACTTACAACATTTGGCGTGAANAACC	1614
Qy	1684	ggctttgctgattgctgattgaaagtgtgaa-----	1715

[illegible]

KEYWORDS	Moraxella catarrhalis.	
SOURCE	Moraxella catarrhalis	
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.	
REFERENCE	1 (bases 1 to 6259)	
AUTHORS	Loosmore,S.M., Sasakii,K., Yang,Y.P. and Klein,M.H.	
TITLE	Recombinant high molecular weight major outer membrane protein of moraxella	
JOURNAL	Patent: WO 0107619-A 12 01-FEB-2001;	
FEATURES	CONNAUGHT LABORATORIES LIMITED (CA)	
source	Location/Qualifiers	
	1..6259	
	/organism="Moraxella catarrhalis"	
	/db_xref="taxon:480"	
BASE COUNT	2067 a 1414 c 1393 g 1385 t	
ORIGIN		
Query Match	35.8%; Score 2484.8; DB 10; Length 6259;	
Best Local Similarity	66.7%; Pred. No. 0;	
Matches 4674; Conservative	0; Mismatches 1067; Indels 1269; Gaps	28;
QY 167	tgtatcggtgcagcgtcaatggcagtgcttatgtctcaaaataatagca-----agatcg	220
DB 2	TGATCGGTGCAAGGCTCAGTGGCAGTGTATGCTCAAAAAAAGATACCAAAACATATCG	61
QY 221	catttgggt-----accacaggcaacaatgacaatgcctcgcgtgacgaatgaagcat	271
DB 62	CAATTGGTGAACAAAACACGACGCTCAGGCACTGCCAAGGCGGACGGTGATCGAG	121
QY 272	ccattctgttagtcttgcgaagccacatgcacatcaagctattgtctatcgtgtgga	331
DB 122	CCATTGCTATTGGTGAATGCTTAACGCACAGGCGGTCAAGCCATCGGCATCGGTAGTA	181
QY 332	gcaaacacagatctctgtaatacagcggcgttaatacagaagcagggtccacgccaagga	391
DB 182	GTAA---TAAACCTGTCAATGGAGGAGCTTTGGATAGATAGTAGTACCGATGCTACGGGTC	238
QY 392	aagaatccatcccatcggtgggtgtagtactggctgaggtgatgctcgtatggcattg	451
DB 239	AAGAGTCCATCGCCATCGGTGGTGTAAAGGCTAGTGGTGATGCTCGATGGCCATCG	298
QY 452	gtagtgaacttatattggataggaatagcactaaacttaataatccaaatgg---tc	508
DB 299	GTAGTGATGACTTACATTTGCTTGATCAGCATGGTAACTCTAAACATCCGAAAGGTACTC	358
QY 509	ttcttagcactcttattcaaaacacatagcattatagcgaatccagactcaaatgggt	568
DB 359	TGATTAACGATCTTTATTAACGGCCATCCAGTATTTAAAGAAATACGAAGCTCAAGGATA	418
QY 569	ctca---gaaataagacgcacagcagaagaagacacgcagcactgctgagtgaggcca	625
DB 419	ATGATGTAAATATAGACGCACAACCCGAAGCGGACACGCCAGTACTGCGAGTGGAGCCA	478
QY 626	tggcatatgcgaagggtaattttccaaagccttggtagacaggtcaacagctgaaggca	685
DB 479	TGTATATGACAGGGTCAATTTTCCAAACGCTTTGGTACACGGGCAACAGCTAANAAGTG	538
QY 686	actatctcttgacgtaggctctaccgccaagcgaaggaataatacaatcgctattg	745
DB 539	CCTATTCCTTGGCAGTGGGTCTTGGCGGCACAGCCGAGGGGCCAATCTACAATCGCTATTG	598
QY 746	gttctaatagcagaactatcaattatggagcactagacccttgggtgcagatactcgaattg	805
DB 599	GTTCTGATCCAACTACTAGCTCGTTGGAGCGCATAGCCCTTGGTGCAGGTAAGTCTGCTC	658
QY 806	atttggattacgtattgctccttaggttatgtctcagatccttaataataataata	865
DB 659	ACTACAGGGCAGTATTCCTTAGCTCAAGGTTCTGTTGTCACACAGAGTATAATAATT	718
QY 866	ataataaagccttatgtaccagaaggttaatgggtcaaacataaaatcgctcaag	922

|||||
Db 3965 AAACGTCCAAAGGGCAGCCAGCAGCACTACAGAGGCTATGTGATGCTGATGCA 4024
Qy 4013 acaaggtcactatgacagttaccgataagaagtactactcaagtcaatgacaaaggtcaag 4072
Db 4025 ATAAAGTCTATCTATGACAGTACCGATTAACAGTACTATCAAGGCCAAAATGATGGCAGAG 4084
Qy 4073 tggacaaaaaagaagttgccaagaagcaaaactgtgcgccaagcccaaaacccagatg 4132
Db 4085 TTGATAAAAACCAAGATTGCCAAGAACAACTGGTCGCCCAAGCCCAACCCAGATG 4144
Qy 4133 gcacattggctcaaatgaatgtcaaatcagttcaattaaacaaagcaagtaaatgatgcca 4192
Db 4145 GCACATTGGCTCAATGATGTCAATCAGTCATTAAACAAAGAACAAAGTAAATGATGCCA 4204
Qy 4193 ataaaagcaaggtcatcaatgaagcaacgcctttatcaaaaggttgaacacgcgcca 4252
Db 4205 ATAAAAGCAAGGCATCAATGAAGACACAGCGCTTTGTTAAAGGACTTGAAAAGCGGTT 4264
Qy 4253 aagacacaaaacaaaacgcgcagtaactgtgggtgatttaaatgcgttgcgcaaa 4312
Db 4265 CTGATTAACAAAACCAAAACCGCGCAGTAACCTGTGGTGATTTAAATGCCGTGGCCAAA 4324
Qy 4313 caccgtgacctttgcaggggatacaggccacacgcgttaaaactggcgagactttga 4372
Db 4325 CACCGTGACCTTTGCAGGGGATACAGGCACAACGGCTAAAAAGCTGGGCGAGACTTTGA 4384
Qy 4373 caataaaggtgggcaaacagacacaaataagctaacgataataacatcgtgtggtag 4432
Db 4385 CCATCAAAAGGTGGGCAACAGACACCAATTAAGCTAACCCGATTAATCACTCGGTGGGTAG 4444
Qy 4433 caggctactgaggttcactgtcaacttgcgaagacccaactcttaacacgctta 4492
Db 4445 CAGTACTGATGGCTTCACTGTCAAACTTGCCTAAAGACCTAACCAATCTTAACAGCGTTA 4504
Qy 4493 atgcaggtggcaccagaattgataaaaaaggcatctctttttagacgcgaacgcgtcaag 4552
Db 4505 ATGCAGGTGGCACAAAATGTATGACAAAAGGCGTGCTTTTGTAGACTCAAGCGGTCAAG 4564
Qy 4553 ccaagcaaacacccctgtgctaagtgccaatggcgctggacacctgggtgcaaacgcataca 4612
Db 4565 CCAAGCAACACACCCCTGTGCTAAGTGCCATGNGGCTGGGCTGGACCTGGGTGGC----- 4614
Qy 4613 gtaacatcgtgagctgttgatgatacgtatgcgtgagcttgaagcagtttaataagaag 4672
Db 4615 ----- 4614
Qy 4673 ttgccaacacggTcaacaaaccttaacacaaaagtaactcaggtgcgtcatattaccctttg 4732
Db 4615 ----- 4614
Qy 4733 tggtaaccgatgccaatggcaagcccatcaatggcaccgatggcaagcccccaaaagcca 4792
Db 4615 ----- 4614
Qy 4793 tcaagggcgcgatggtataactactatcaagccaacccaacgcggtacactgtggacaag 4852
Db 4615 ----- 4614
Qy 4853 atggcaagcccatcaaccgatgcggaacaaacttgccaatctggcagctcatggcaacccc 4912
Db 4615 ----- 4614
Qy 4913 ttgatcaggtcatcaagtgtggcaagccttagcgcgcaactcagatgccatcacccctaa 4972
Db 4615 ----- 4614
Qy 4973 ccaacatcaagtccaactttgccacaaattgacacacaaacacaggtlaatgccaatgacag 5032
Db 4615 ----- 4614
Qy 5033 ggcaagcccaagtcctgccagcctatcagcagcacagcaaaagtaatgctgcagtgtoa 5092

Db 4615 ----- 4614
Qy 5093 aagatgtgctaaatgtaggctttaacttgagaccgaatcaccaatcaagtggaactttgtca 5152
Db 4615 ----- 4614
Qy 5153 aagcctatgataccgtcaactttgtcaatggtacaggtgcccagacatcacaaagcgtgcgta 5212
Db 4615 ----- 4614
Qy 5213 gtgctgatggcacgatgagtaaacatcacccgtcaacacgcgcttagcagcgaccgatgatg 5272
Db 4615 ----- 4614
Qy 5273 atggcaatgtgcttatcaaaagcaaaagatggttaagtctctacaaagcagacgcctcatgc 5332
Db 4615 ----- 4614
Qy 5333 caaacggctcactaaaagcaggcaaaatcagccagtgatgccccaaactccaaactggctctaa 5392
Db 4615 ----- 4614
Qy 5393 gccttgtaaccccaatgctgttaaaagcagtcacaggcgatgcgtgctcttaataact 5452
Db 4615 ----- 4614
Qy 5453 tatcaaaagcgtgatttaaatccaaagatggtacaactactaccacagtaagctctgatg 5512
Db 4615 ----- 4614
Qy 5513 gcatacagtatccaagggcaaaagataaacagcagcatcacccctaagcaaaagatgggtgaatg 5572
Db 4615 ----- 4614
Qy 5573 taggcgttaaggtcatcagcaatgtgggtaaaaggcacaaaaagacaccgcgctgcgaatg 5632
Db 4615 ----- 4614
Qy 5633 tacaacagttaaacgaagtacgcgaactgttgggtccttgggttaagtctggttaagtatacag 5692
Db 4667 TACAACAGTTTAACCAAGTACGCAACTTGTGGTCTTGGTAACTGCTGCTAATGATAACG 4726
Qy 5693 ctgacgcgaatcagtgtaaacattgcgcgacatcaaaaaagacccaattcaggttcatcat 5752
Db 4727 CTGACGGCAATCAGGTAAACATTGCCGACATCAAAAAGACCCCAATTCAGGTTTCATCAT 4786
Qy 5753 cttaacgcgactgtcatcaaaagcaggcacggttacttggcggttaaaaggttaataacgatacag 5812
Db 4787 CTAAACGGCACTGTCTATCAAAAGCAGGCACGGTACTTGGCGGTAAAGGTAATAACGATACCG 4846
Qy 5813 aaaaaacttgccaactgtggtgtacaagtggtgtgataaaagcagcgaacgcgcaacgcgctg 5872
Db 4847 AAAAACTTGCCCACTGGGTGTATACAAGTGGGCTGTGATTAAGACCGCAACGGTACACGGCG 4906
Qy 5873 atttaagcaatgtttgggtcaaaaccccaaaaagatggcagcaaaaaagccctgctgcgca 5932
Db 4907 ATTTAAGCAATGTTTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGCTCGCCA 4966
Qy 5933 cttaataacgcgacggttcagacccaactattgacccaacccccccagagcattgaca 5992
Db 4967 CTATTAACGCCCGCAGGTACAGCAACTATTGTACCAACAAACCCCGCAGAAAGCATTTGACA 5026
Qy 5993 gaataaatgaacaaaggtatccgcttcttccatgtcaacgatggcaatcaagagcctgtgag 6052
Db 5027 GAATAAATGAACAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCCTGTGG 5086
Qy 6053 tacaagggcgtaacggcatttgactcaagtgcctcagggaagcactcagtggtgcgataagtt 6112
Db 5087 TACAAGGGCTTAACGCAATTGACTCAAGTGCCTCAGGCAAGCACTCAGTGGGCGATAGGTT 5146
Qy 6113 tccaggccaagcagatggtgaagcccgcttgccatagcagacaaaaccccaagcagga 6172
Db 5147 TCCAGGCCAAGGCAGATGATGTTGAAGCCCGCTTGCCTATAGGCACACNACCCCAAGCAGGCA 5206


```
QY 6173 accaatccatcgccatcggtgtataacgcacaaagccacggcgcatcaatccatcgccatcg 6232
Db 5207 ACCAATCCATCGCATCGGTGATAACGCACAAAGCCACGGCGCATCAATCCATCGCCATCG 5266
QY 6233 gtacaggcaatgtggttaaacaggttaagcactctgtgcccactcgccacgaactgtta 6292
Db 5267 GTACAGGCAATGTGTTAGCAGGTAAAGCAGCTCTGTGCGCATCGGCGACCCAGGACTGTA 5326
QY 6293 aggcgtataacagttacagtggtggttaataacaacacagttatcgtatgccactcagaccg 6352
Db 5327 AGGCTGATAACAGTTACAGTGTGGTAAATAACAACACAGTTTACCGATGCCACTCAACACG 5386
QY 6353 atgtctttgtgtgggcaataacatacaccgtgaccgaaagtactgtgtgctttaggtt 6412
Db 5387 ATGTCTTTGTGTGGGCAATAACATCACCTGACCGTGAACCGAAAGTAACCTCGGTTGCCTTAGGTT 5446
QY 6413 caaatctgcacatcagtgaggcacacacagcagccacacagccaaataatctgacggca 6472
Db 5447 CAACCTCTGCCATCAGTGCAGGACACACACGACGACACAAAGCCCAAAAATCTGACGGCA 5506
QY 6473 cagcaggtatacaacacacacagcaggtgcaacaggtacgggttaagggctttgclggacaaa 6532
Db 5507 CAGCAGGTACACACACACAGCAGGTGCAACCGTACCGGTTAAAGGCTTTGCTGGACAAA 5566
QY 6533 cggcgggtgtgctgcgttcgtgtggtgctcaggtgctgaacgcttatccaaatgtgg 6592
Db 5567 CGCGGGTGTGTGGGCTCTCGGTGGTGCCTCAGGTGCTGAACGCGGTATCCAAAATGTGG 5626
QY 6593 cagcaggtgaggtcagtgccacacagcagcagcagcagcagcagcagcagcagcagcagcag 6652
Db 5627 CAGCAGGTGAGGTGAGTGCAGGACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5686
QY 6653 ccaccccaagcattgccaacgcacacaaatgagcttgaccatcgatccaccccaaacgaaa 6712
Db 5687 CCACCCAAAACATTGCCAACCGCAACCAATGAGCTTGACCATCGTATCCACCAAAACGAAA 5746
QY 6713 ataaagccaatgagggatttcacagcagcagcagcagcagcagcagcagcagcagcagcagc 6772
Db 5747 ATAAGGCCAATGACAGGATTTCATCAGCGATGGCGATGGCGTCCATGCCACAAAGCCTTACA 5806
QY 6773 ttctcgcagatccatggttacgggggtatttgcaccccaacagcagcagcagcagcagcagcagc 6832
Db 5807 TTCTTGCAGATCAATGTTTACCGGGGTATTTGCCACCCACACAGCGTCAAGGTGCGGTGG 5866
QY 6833 cagtggaactgcgaagctgcggataatggtcaatgggttatataaatcaatggttcag 6892
Db 5867 CAGTGGGACTGTCGAGCTGTCGATATGTCATATGGGTATTTAAATCAATGTTTCAG 5926
QY 6893 ccgatacccaagccatgtaggggcagcagcagcagcagcagcagcagcagcagcagcagcagc 6942
Db 5927 CCGATACCCAAAGCCATGTAGGGGCGCAGTTGGTGAGGTTTTCACCTT 5976

RESULT 10
AX079957
LOCUS AX079957 3135 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 45 from Patent WO0107619.
ACCESSION AX079957
VERSION AX079957.1 GI:13159465
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 3135)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer membrane protein of
Moraxella
JOURNAL Patent: WO 0107619-A 45 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
Location/Qualifiers
source 1..3135
```

/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

BASE COUNT 1011 a 746 c 741 g 637 t

ORIGIN

Query Match 25.28; Score 1752.4; DB 10; Length 3135;
Best Local Similarity 71.18; Pred. No. 0;
Matches 2954; Conservative 0; Mismatches 166; Indels 1032; Gaps 7;

QY 2803 ggtacggttaccttggcattacacacaaagcggtcttaaacggcgccagcagcact 2862

Db 1 GGTACGGTTACCTTGGCATTACACCAAGCGGTCTTAAAGCCGCGCAAGCACC--- 57

QY 2863 ctaaacaaagatgcttctctattataaacccccgcctagtacgacaaatccaagtgg 2922

Db 58 CTRAAACGACGGTGGCTTGTCTATTATAAACCCCTGTAGTACGAACAAATCCAAGTCGGT 117

QY 2923 gctgatggcgtgaagttggcaaggt---tgaataagggtaattcaagcactggcattgat 2979

Db 118 GCTGATGGCGTGAAGTTGGCAAGGTTAATAATAATAGTGTGTAGGTGCTGGCATTGAT 177

QY 2980 ggcacaagcgtatcaccaaaagatcaaatggctttactgggctaatggcctcacttgat 3039

Db 178 GGCACAACTCGCATACCAAGATGAATGGCTTTACTGGGACTAATGGCTCACTTGTAT 237

QY 3040 accacacacccccacctaaccacaaagacaaagcttaagtgaggtgaaattgaaattaccaac 3099

Db 238 AAAAGCAAAACCCACCTTAAGCAAAAGAC----- 264

QY 3100 actggccttaacgcaggtggttaaaagattaccacaaattcaatcaggtgataattaccacaa 3159

Db 265 ---GGCATTAACGCGAGGTGTTAAAGATTACCAACAAATTCATCAGGTGAGATTGCCCAA 321

QY 3160 aacagaatgatgctgacagcgctcggtttatgatattaaaaacgcgaacttgaagc 3219

Db 322 AACAGCATGATGCTGTGACAGGGCGCAAGATTATGATTTAAAAACCGAACTTGAANAAC 381

QY 3220 aaatcaacagtgctgttaaaacagcagcaaaactcattacagaaattctcagtagcagat 3279

Db 382 AAATCAGCAGTACTTGCCTGCAAAACAGCAAAATCATTACAGAAATCTCAGTAGCAGAT 441

QY 3280 gaacaaggttaacttaaggttagtaaaccttactccagttatgacacotcaaaagacc 3339

Db 442 GAACAAGGTATAACTTTACGGTTAGTAACCTTTACTCCAGTTATGACACCTCAAGACC 501

QY 3340 tctgatctalcaccttgcaggtgaaacggcattaccaccaaaggtataaataaagggtg 3399

Db 502 TCTGATGTCATCAGCTTTGACAGGTGAACCGCAATACCACCAAGGTAAATAAAGGTGTG 561

QY 3400 gtgcgtgtggcattgacacaaacaaaggtttaaccacgcctaaagctgacccgtgggtaat 3459

Db 562 GTGCGTGTGGCATTGACCAAAACCAAGGCTTAACCCAGCCCTAAGCTGACCGTGGTAAAT 621

QY 3460 aataatggcaaggcattgtcattgacagtaaaagatgggtcaaaataccatcacaggacta 3519

Db 622 AATAATGGCAAGGCATTTGATTGACAGCCAAATGGTCAAAAATACCATCACAGGACTA 681

QY 3520 agcaacactctagtaagtgtaccaatga-----tggtgcagacacgcagcactaagccaa 3573

Db 682 AGCAACACTCTAGCTAATGTTTACCATGATAAAGGTAGCTGACGACACACAGAGGC 741

QY 3574 gggccttgccaatgacacgcagcaaaaccccgctgcccgccagcattggtgattgtctaaacga 3633

Db 742 AATATAATCAAGACGAAGACAAAACCCGTCGCCGCGCAGCATTTGTATGTGCTTAAGCGCA 801

QY 3634 ggccttaacttgcaggaatgtgtaagcggttgactcttctccacttatgacactgtt 3693

Db 802 GGCCTTAACCTGCAAGCAATGTTGAAGCGGTTGACTTTGTCTCCACTTATGACACCGTC 861

QY 3694 gactttatcatggcaatgccaccacgcctaaagtgacctatgatgacacaaagcaacc 3753

Db 862 AACTTTCCGATGCAATGCCACCACCGCTAAGGTGACCTATGATGACACAAAGCAACCC 921


```
Db 2101 AGCAAAAGCCCTGCTCGCACTTATAACGCCGCGAGTCAGACCAACTATTGACCAAC 2160
QY 5971 aacccgcagaagccattgacagaataaataaacaaggtatccgcctctccatgtcaac 6030
Db 2161 AACCCGCGAAGCCATTGACAGATTAATGACACAGGTATCCGCTCTTCATGTCAAC 2220
QY 6031 gatggcaatcaagagcctgtggtacaaagggcgtaacggcattgactcaagtgcctcaggc 6090
Db 2221 GATGGCAATCAAGAGCCTGTGTGTAAGGGCGTAACGGCATTCAGTCAAGTGCCTCAGGC 2280
QY 6091 aacactcaatgcataggtttccaggcccaagcagatggtgaagccgcggttgccata 6150
Db 2281 AAGCACTCACTGGCGATAGGTTTCCAGGCCAAGCGATGTTGAAGCCGCGGTTCGCATA 2340
QY 6151 ggcagacaacccaagcaggaaccaatccatcgccatcggtgataacgcacaaagccaag 6210
Db 2341 GGCAGACAACCCANGCAGCAACCAATCCATGCCATCGGTGATACGGCACAAAGCCAGC 2400
QY 6211 ggcgataatccatcgccatcggtgacagggcaatgtggtacacaggtgaagcactctgtgcc 6270
Db 2401 GGGCATCAATCCATCGGCATCGGTACAGCAATGTGTTAGCGTAAGCACTCTGTGTGCC 2460
QY 6271 atcgggacccaagcactgttaaggctgataaacagtttaacagttgggttaatacaaccag 6330
Db 2461 ATCGGCGACCCAGCACTGTTAAGGCTGATAACAGTTTACAGTTGCGGTATTAACCAACCAG 2520
QY 6331 ttcatgatccactcagacccgatgtcttttgggtggtggcaataacatcacctgacccgaa 6390
Db 2521 TTACCAGTCCCACTCAACCCGATGTCTTTGGTGTGGCAATACATCACCCGTCACCGAA 2580
QY 6391 agtaactcgtgtccttagatgtcaaacctcgtccatcagtcagggcacacacgcagggcaca 6450
Db 2581 AGTAACTCGGTTGCTTAGTTTAAACTCTGCCATCAGTCAGTCAGGCACACACGAGGCACA 2640
QY 6451 caagccaaaatctgacgcacagcaggtacaaaccaacacagcaggtgcaacaggtacg 6510
Db 2641 CAGCCAAAATAATCTGACGCGACAGCAGGTGACCAACCAACCAACAGCAGGTGCAACCGTACG 2700
QY 6511 gtaaaagccttctggtgacaaaagcggtgtgttgggtgcgtctccgtgggtgcctcaggtgct 6570
Db 2701 GTTAAAGCCTTTGCTGACAAACGGCGGTTGGTGGGTCTCGGTGGTGGCTCAGTGGCT 2760
QY 6571 gaacgcgataccaaaatgtyggcagcaggttgagtgatgctgccaccagcaccgatggggtc 6630
Db 2761 GAACGCGTATCCAAAATGTGGCAGCAGGTGAGTCACTGCTCAGTCCACCAACCGATGCGGTC 2820
QY 6631 aatggttagcagttgtacaagccaccaccccaagcagcatgtgccaacgaaccaatgagcttgac 6690
Db 2821 AATGGTAGCCAGTTGTACAAGGCCACCCAAAGCATTGCCCAACGCAACCAATGAGCTTGAC 2880
QY 6691 catcgtatccaccacaaaataaagcccaatgacggggtatttcacagcgatggcgatg 6750
Db 2881 CATCGTATCCACCAACAGCAATAGGCCAATGCGAGGATTTCAATCAGCGATGGCGATG 2940
QY 6751 gcttcaatgcacaagcctacattccttggcagatcccatggttaccgggggtattgcccacc 6810
Db 2941 CGCTTCATGTCACAAGCCTACATTCCTTGGCAGATCCATGTTTACCGGGGGTATTGCCACC 3000
QY 6811 cacaacggtcaagtgccgtgagcagttggactatcgaagctgtcgataatggtcaatgg 6870
Db 3001 CACACGGTCAAGTGGCGGTGGAGTGGGACTGTCGAAGCTGTCGAAGTATGGTCAATGG 3060
QY 6871 gtatttaaatcaatggttcagccgaaccacgaagggccatgtaggggcgaggttgggtgca 6930
Db 3061 GTATTTAAATCAATGTTTCAGCGCATACCCAGGCCATGTAGGGGGCGCGATGTGGTGCA 3120
QY 6931 ggttttcaacttt 6942
Db 3121 GGTTTTCACTTT 3132
```

RESULT 11

```
AX079959
LOCUS 2448 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 47 from Patent WO0107619.
ACCESSION AX079959
VERSION AX079959.1 GI:13159466
KEYWORDS Moraxella catarrhalis.
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
REFERENCE 1 (bases 1 to 2448)
AUTHORS Loomore,S.M., Sasaki,K., Yang,Y.P. and Klein,M.H.
TITLE Recombinant high molecular weight major outer Membrane protein of moraxella
JOURNAL Patent: WO 0107619-A 47 01-FEB-2001;
CONNAUGHT LABORATORIES LIMITED (CA)
FEATURES
source 1..2448
/organism="Moraxella catarrhalis"
BASE COUNT 826 a 569 c 556 g 497 t
ORIGIN
Query Match 20.5%; Score 1421.8; DB 10; Length 2448;
Best Local Similarity 87.2%; Pred. No. 0;
Matches 1648; Conservative 0; Mismatches 187; Indels 54; Gaps 6;
QY 2803 ggtacggttacccttgccattaacacccaaagcggtcttaaaagcggcgacagcaccact 2862
Db 1 GGTACGGTTCCTTTGGCATTACACCAAGCGGTCTTAAAGCCGCGCAAGACC--- 57
QY 2863 ctaaaacgaagtcgttctctattaaaaccccgctgtagtaacgacaaatccaaagtcggt 2922
Db 58 CTAACCCGCGTGGCTGTCTATTAAAACCCCACTGGTAGCGAACAATCCAAGTCGGT 117
QY 2923 gctgtagcggtgaagtttgccaaagt---tgataagggtaattcaagcactggcattgat 2979
Db 118 GCTGATGCGGTGAAGTTTGCAAGGTTAATAATAATGCTGTGTAGGTGCTTGGCATTTGAT 177
QY 2980 ggcacaagcgtatcaccaagatcaaatggcttcttactggggtggttaagtcctactgat 3039
Db 178 GGCACAACCTCGCATACCAAGATGAATTTGGCTTTTACTTGGGACTAATGGCTCACTTGAT 237
QY 3040 accacaaaacccccacctaaaccaagacaaagcttaaaagtggtgaggtgaaattaccaac 3099
Db 238 AAAAGCAAAACCCCACTTAAGCAAGAC----- 264
QY 3100 actggcattacgcaggtggttaaaagattaccaacattcaatcaggtgatattacccaa 3159
Db 265 ---GGCATTTAACGCAGGTGGTAAAAAGATTACCAACATTTCAATCAGGTGAGATTGCCAA 321
QY 3160 aacagcaatgctgtgacagcggtcggtgttatgtatttaaaacccaactgaaagc 3219
Db 322 AACAGCATGATGCTGTGACAGCGCGCAAGATTATGATTTAAAACCGCAACTTGAANAAC 381
QY 3220 aaaaatcaacagtgctgtaaaacagcaaaaactcattacacgaattctcagtagcagat 3279
Db 382 AAAATCAGCAGTACTGCCAAAACAGCAAAACATCAATACAGAAATTCCTCAGTAGCAGAT 441
QY 3280 gaacaaggttaateactttacggttagtaaaccttactccagttatgacacctcaaaagacc 3339
Db 442 GAACAAGGTATAATCTTTACCGTTTAGTAACCCCTTACTCCAGTTTATGACACCTCAAGACC 501
QY 3340 tctgatatcatcaccttgcaggtgaaagcgcttaccaccaaggttaataaaggtgtg 3399
Db 502 TCTGATGTCATCACTTTGAGGTGAAACGGCATTTACCACCAAGGTAATAATAAGGTGTG 561
QY 3400 gtgcgtgtgggcatttgaccaaaacaaaggtcttaaccacgcctaaagctgacgtgggtaat 3459
Db 562 GTGCGTGTGGCATTGACCAAAACCAAGGCTTAACCAAGCCCTAAGCTGACCGTGGGTAAT 621
QY 3460 aataatggcaaggcattgttcattgacagtaaaagtggtcaaaataccatcacaggacta 3519
```


Thu Sep 13 14:17:47 2001

```

Db 1699 TTTGTAGACTCAAGCGTCAAGCCAAAGCAACACCCCTGCTTAAGTGCCAAATGGGCTG 1758
QY 4591 gacctgggtggcaaacgcatcaatcaatcgctgagctgtgtgatgataacgagtcgctg 4650
Db 1759 GACCTGGGTGGCAAGTGCATCAGTAATGTGGCAAGGACCAAAAGATACCGACGCTGCC 1818
QY 4651 aactttaagcagcttaataagaagtgtgcaa 4679
Db 1819 AATGTACACAGCTTAACGAGTACGCAA 1847

RESULT 12
PMU277635 3900 bp DNA BCT 21-OCT-2000
LOCUS Pasteurella multocida mapA gene.
ACCESSION AJ277635
VERSION AJ277635.1 GI:10945096
KEYWORDS MapA protein.
SOURCE Pasteurella multocida.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
          Pasteurella.
REFERENCE 1 (bases 1 to 3900)
AUTHORS Henderson, I.R., Nataro, J.P., Cappello, R. and Stein, C.
TITLE Evolutionary origins of the autotransporter proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3900)
AUTHORS Henderson, I.R.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) Henderson I.R., Center for Vaccine
          Development, University of Maryland School of Medicine, 685 W.
          Baltimore St, MD 21202, USA

FEATURES             Location/Qualifiers
     source            1..3900
                        /organism="Pasteurella multocida"
                        /strain="PM70"
                        /db_xref="taxon:747"
     gene              1..3900
                        /gene="mapA"
     CDS                1..3900
                        /gene="mapA"
                        /function="putative autotransporter protein"
                        /transl_table=11
                        /product="MapA protein"
                        /protein_id="CAC14202.1"
                        /db_xref="GI:10945097"
                        /translation="MVKIYRTLMNAATQSWVYSELAKAGKSGKSLVNSVSGFS
FTLIAASVVLGSGOVNAEVTGNTGVSGDDKYCFYNASSQSVICGDATTTKTDKDNK
PAKSVVIGFATNDGETNVALGAKSKSKAASIAIGDNKALDNOAIAIGONATNSD
WDISIGROAGAEQTEVSAEGRNIAIGDGLKRGKGVNNIALGTSAGDLAGTHNVLM
GTYYNADAVRSALTAGSTTKKEINAKETDNKTYIEASNVVALGTRALATQLAARVAG
QQAFAFGNSVAVNGTKASGQTAIATGSHNAHATGSSIAIGTVNGNDRTARTLASDN
FTIAMLGSAQATKDAIAVGRNKAAGINTVSYGNAGVTKTADAQSDPNKLEPAT
DAVFIGNKAGYKSNQRMQVSLGDSGVEVGTENVITGNSAGKNTKGNNTVSLKESVSLK
CONVEHNPFAALLLEAGQNIKGSQNIAGKHGRSADPNTKLINNTTISLCKESVSLK
NFGTAKGNKAKTDGLASTAIGRNAEAVGGETNANTIAIGDSASADAGIAVLGTQAQKS
LTVDGKYGAYSIVIGTEAKIAQAAPAGKNENPKDAIAGTKAEAHYASTIALGFG
AKSDTKAQAVSIGYNSNAKQYATAFGEAKTENAGSSIAFGTKAQTTRASASTAIG
GAETFGDGGQALDSDAVALGREAKQNALAFYKAVADHKDAVALGAGNAETAAAE
GTNEATVNEFYSGPAGIKPTATVSVGKKDAERTITNVAAGRIDKTDIDAINGSLYL
ALNALGVNGNTLVTNLVCGDAALVKEGDEAGTLTMSNIGGTGKGTIHDAIPAVNTAK
ASKTTPVKEGDMITVTEEAADGSRITYTATKDKVDFSDVAVAGGTFAQGEVKNIIIGDIT
TKIDNTPSISKIGIDNGKQVTVNONGNIAKDSKDAVNGOLFQAGGCVKNIIIGDIT
YNPTEGEANTNIGGTGASTIDEAKVNTTAKAKTEVVOGENIVTVSAPGANGTV
YTVATKEVTFDKTTVGSVVDKNTDITJLSNKTIGDNGNPAKNGRAASESOLNATOT
NIATLGGNAONTNGVAMTDIGGTGKNNDIAIKASRNEVKQGNKMNVTPTTCANGQ
TIVYVATADKVAFDVKVGGITIDATTNKISGIAKGDISENTDANGVSQGLYELOQKI
AKSGNYNTLNNRINKVDRLRAGIAGANAAGLPQAYIPGKSMVAVAAGTYKQNAI
ALGMSRIDNGHVKIILKTGNTNSRGDFGASIGAGYQW"
1320 a 688 c 934 g 958 t

BASE COUNT
ORIGIN
```


us-09-361-619-10.rge

Thu Sep 13 14:17:47 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 13:39:40 ; Search time 1359.14 Seconds
(without alignments)
3183.672 Million cell updates/sec

Title: US-09-361-619-10

Perfect score: 6942

Sequence: 1 atgaatacatctataagt.....ttgtgcaggtttcacttt 6942

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDSI/gcgdata/geneseq/geneseq/NAL1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq/NAL1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq/NAL1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq/NAL1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq/NAL1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq/NAL1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq/NAL1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq/NAL1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq/NAL1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq/NAL1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq/NAL1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq/NAL1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq/NAL1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq/NAL1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq/NAL1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq/NAL1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq/NAL1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq/NAL1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq/NAL1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq/NAL1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6942	100.0	6942	22	AAF59105 M. catarrhalis les
2	2939.2	42.3	6159	22	AAF59104 M. catarrhalis str
3	2852.4	41.1	62909	22	AAF28545 Genomic fragment #
4	2590	37.3	6144	22	AAF59103 M. catarrhalis str
5	2590	37.3	6972	22	AAF59102 M. catarrhalis str
6	2580.6	37.2	6971	17	AAF38740 Moraxella outer me
7	2580.6	37.2	6973	22	AAF59100 M. catarrhalis str
8	2485.8	35.8	5979	22	AAF59101 M. catarrhalis str
9	2484.8	35.8	6259	22	AAF59106 M. catarrhalis M56
10	147.8	2.1	3030	21	AAF59106 Haemophilus influe
11	146.2	2.1	3036	21	AAF59106 Haemophilus influe

12	141.4	2.0	3354	21	AAF92495	Haemophilus influe
13	132.6	1.9	3342	21	AAF92498	Haemophilus influe
14	116.6	1.7	720	22	AAF59129	M. catarrhalis str
15	110.6	1.6	936	22	AAF58252	Oligonucleotide D1
16	110.6	1.6	936	22	AAF58254	Oligonucleotide D1
17	110.6	1.6	936	22	AAF58257	Oligonucleotide D1
18	110.6	1.6	936	22	AAF58259	Oligonucleotide D2
19	110.6	1.6	936	22	AAF58262	Oligonucleotide D2
20	110.6	1.6	938	22	AAF58255	Oligonucleotide D1
21	110.2	1.6	936	22	AAF58252	Oligonucleotide D1
22	110.2	1.6	936	22	AAF58254	Oligonucleotide D1
23	110.2	1.6	936	22	AAF58257	Oligonucleotide D1
24	110.2	1.6	936	22	AAF58259	Oligonucleotide D2
25	110.2	1.6	936	22	AAF58262	Oligonucleotide D2
26	83.6	1.2	2037	17	AAF41477	Haemophilus adhesi
27	82.8	1.2	3294	17	AAF41475	Haemophilus adhesi
28	82	1.2	2079	21	AAF92494	Haemophilus influe
29	80.2	1.2	7486	19	AAF22837	Haemophilus paraga
30	78.6	1.1	8930	19	AAF22834	Haemophilus influe
31	76.6	1.1	1830121	17	AAF42063	Haemophilus influe
32	76.4	1.1	3300	21	AAF92497	Haemophilus influe
33	76.4	1.1	7253	21	AAF92499	Haemophilus influe
34	76.4	1.1	7291	17	AAF41476	Haemophilus adhesi
35	66	1.0	1757	21	AAF92500	Haemophilus influe
36	59.6	0.9	2700	21	AAF70220	Plasmodium falcipa
37	58.4	0.8	60	17	AAF38741	Moraxella outer me
38	58.4	0.8	1797	20	AAF85794	DNA encoding a sur
39	56.8	0.8	1797	20	AAF85790	DNA encoding a sur
40	55.8	0.8	1779	20	AAF99125	DNA encoding N. me
41	55.2	0.8	1770	20	AAF85797	DNA encoding a sur
42	55.2	0.8	1776	20	AAF99124	DNA encoding N. me
43	55.2	0.8	1776	20	AAF85798	DNA encoding a sur
44	55.2	0.8	1776	20	AAF85798	DNA encoding a sur
45	55.2	0.8	1776	20	AAF85793	DNA encoding a sur

ALIGNMENTS

RESULT 1

AAF59105
ID AAF59105 standard; DNA; 6942 BP.
XX AAF59105;
XX AC
XX 24-APR-2001 (first entry)
XX M. catarrhalis les1 200kDa gene SEQ ID NO:10.
DE
XX
KW Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX WPI; 2001-159722/16.
XX P-PSDB; AAB69136.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis

Claim 1: Fig 5A-Y: 247pp; English.

The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as immunogenic compositions and vaccines to protect against *M. catarrhalis* infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of *M. catarrhalis*. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the *Escherichia coli* host. The present sequence represents the *M. catarrhalis* les1 200kDa gene, which is given in the exemplification of the present invention.

Sequence 6942 BP; 2286 A; 1621 C; 1564 G; 1471 T; 0 other;

Query Match	100.0%;	Score 6942;	DB 22;	Length 6942;
Best Local Similarity	100.0%;	Pred. NO. 0;		
	0.0;	Microatches	0.0;	Indels 0;
		Gaps	0;	

Qy	1	atgaatcacatctataaagtcattcttaacaaagccacaggcacatttatggcgtggca	60
Db	1	atgaatcacatctataaagtcattcttaacaaagccacaggcacatttatggcgtggca	60
Qy	61	gagtgcgcaaatccccacagcgaggagtagcagtagtacgcgacagacagctggcgagc	120
Db	61	gagtgcgcaaatccccacagcgaggagtagcagtagtacgcgacagacagctggcgagc	120
Qy	121	tctctgtcatccgcctgactcgtgtgtgcacagctcgctactctagtcggtgcgacg	180
Db	121	tctctgtcatccgcctgactcgtgtgtgcacagctcgctactctagtcggtgcgacg	180
Qy	181	ctcaatggcagtgcttgcataaataatagaagatgcatttggaccacaggcaac	240
Db	181	ctcaatggcagtgcttgcataaataatagaagatgcatttggaccacaggcaac	240
Qy	241	aatgacaatgcctcggctagcaatgaagcatccattgctattggtagcttgcgaagca	300
Db	241	aatgacaatgcctcggctagcaatgaagcatccattgctattggtagcttgcgaagca	300
Qy	301	catgccaatcaagctattgctatcggtagtgtagcaaacagagatcctctaatcaagcgct	360
Db	301	catgccaatcaagctattgctatcggtagtgtagcaaacagagatcctctaatcaagcgct	360
Qy	361	aatcagaagcgaggttccacgcacaaaggttaaaagtcctatcgccatcgggttgtagtga	420
Db	361	aatcagaagcgaggttccacgcacaaaggttaaaagtcctatcgccatcgggttgtagtga	420
Qy	421	ctggctgagggtagtgctcctcgatgccattggtagtgtagtacttatattggataggaaat	480
Db	421	ctggctgagggtagtgctcctcgatgccattggtagtgtagtacttatattggataggaaat	480
Qy	481	agcaactaacctctaaatatccaaatgctctcttagcacctcttatccaaacccatcacgta	540
Db	481	agcaactaacctctaaatatccaaatgctctcttagcacctcttatccaaacccatcacgta	540
Qy	541	ttacgcgaataacagagactcaaatggttcttcagaaaaatagacgcacagcagcagaagga	600
Db	541	ttacgcgaataacagagactcaaatggttcttcagaaaaatagacgcacagcagcagaagga	600
Qy	601	cacgccagtactcagctgggagcgcataggcattatgcgaagggtcattttgccaacgccttt	660
Db	601	cacgccagtactcagctgggagcgcataggcattatgcgaagggtcattttgccaacgccttt	660
Qy	661	ggatcacggctcaacagctgaagcacaacttctctggcagtaggtcttaccgccaaaagcc	720
Db	661	ggatcacggctcaacagctgaagcacaacttctctggcagtaggtcttaccgccaaaagcc	720

Db 1801 ||||| aagatcagtaagttaaaagatgcaacgagacgataccgagtgacgtgacgtacattataacagcgtt 1860
QY 1861 aaacaagttcaacaagacccgacggtgcccatacaagcttctctattcgtgatgaaaaa 1920
Db 1861 aaacagttccaaacagacccgacggtgcccatacaagcttctctattcgtgatgaaaa 1920
QY 1921 ggtcaggaattacagattagtaacttgatttctaagttaacaccccaataacctttgag 1980
Db 1921 ggtcaggaattacagattagtaacttgatttctaagttaacaccccaataacctttgag 1980
QY 1981 accatacctttcaggtgaaacggtacatcagttatcagcaaatgacatagccaaggttaa 2040
Db 1981 accatacctttcaggtgaaacggtacatcagttatcagcaaatgacatagccaaggttaa 2040
QY 2041 gtcaaaagtgtgattgacccaatacaatgggtctcacacgcttaagctgacccgtgggtagc 2100
Db 2041 gtcaaaagtgtgattgacccaatacaatgggtctcacacgcttaagctgacccgtgggtagc 2100
QY 2101 gataaagatggttaaaactcaattgggtattgagcaagtggtgagcgttaacgacacaaa 2160
Db 2101 gataaagatggttaaaactcaattgggtattgagcaagtggtgagcgttaacgacacaaa 2160
QY 2161 acatcattagagattgtcccacaaactgcttagcattaccatgacgtggtgacgc 2220
Db 2161 acatcattagagattgtcccacaaactgcttagcattaccatgacgtggtgacgc 2220
QY 2221 accacagaacagggcaatacaatcacacgagcagagacaaaatccaaagcccgagctac 2280
Db 2221 accacagaacagggcaatacaatcacacgagcagagacaaaatccaaagcccgagctac 2280
QY 2281 ggtgatatataacacagctttaacctaaaaataataagcaactcgttgctttgtc 2340
Db 2281 ggtgatatataacacagctttaacctaaaaataataagcaactcgttgctttgtc 2340
QY 2341 tcaactataacactgttgactttatogatggcaatgccaccaccgcttaaggttaacttac 2400
Db 2341 tcaactataacactgttgactttatogatggcaatgccaccaccgcttaaggttaacttac 2400
QY 2401 gatgaacccaatacaaacccagtaagtaacttatgatctcaatgltgagtgaagaaacatt 2460
Db 2401 gatgaacccaatacaaacccagtaagtaacttatgatctcaatgltgagtgaagaaacatt 2460
QY 2461 gaactcacaggcgataatggtcagagacaaaacaaaattggcgtcaaaaacccacactgacc 2520
Db 2461 gaactcacaggcgataatggtcagagacaaaacaaaattggcgtcaaaaacccacactgacc 2520
QY 2521 acaacaaatgctaagtgtaaagcaacaaacttagtaaccaaccgataacgctgctttgtt 2580
Db 2521 acaacaaatgctaagtgtgtaaagcaacaaacttagtaaccaaccgataacgctgctttgtt 2580
QY 2581 aacgccaagacalcgcccgaatacttaaacaccctagccaaggaatactcacaccacaaa 2640
Db 2581 aacgccaagacalcgcccgaatacttaaacaccctagccaaggaatactcacaccacaaa 2640
QY 2641 ggcacagcagacacccgctctacaacactttaagtcataaaaagacggtgcaactgtgac 2700
Db 2641 ggcacagcagacacccgctctacaacactttaagtcataaaaagacggtgcaactgtgac 2700
QY 2701 gaaacataccgtgggtaaagatggtacacaaaagcgcaagaccgtcaacactctaaaa 2760
Db 2701 gaaacataccgtgggtaaagatggtacacaaaagcgcaagaccgtcaacactctaaaa 2760
QY 2761 ctcaaaagtgaacacggtctcaacggttgctaccaataaagatggttacgtttacgtttggc 2820
Db 2761 ctcaaaagtgaacacggtctcaacggttgctaccaataaagatggttacgtttacgtttggc 2820
QY 2821 attaacacccaaagcgtcttaagccgagcagcaccactctaaacaaagatggcttg 2880
Db 2821 attaacacccaaagcgtcttaagccgagcagcaccactctaaacaaagatggcttg 2880
QY 2881 tctattaaaaaccccgctagttaacgaacaaaatccaagtcggtgctgtagcgtgaagt 2940
Db |||||

Db 2881 tctattaaaaaccccgctagtaacgaacaaaatccaagtcggtgctgtagcgtgaagt 2940
QY 2941 gccaaagttgataagggtaattcaagcactggcattgatggcacaagccgtatcaccaaa 3000
Db 2941 gccaaagttgataagggtaattcaagcactggcattgatggcacaagccgtatcaccaaa 3000
QY 3001 gatcaaatggcttacttgctgggttaattggctcactgtatcaccacaaaacccacctaacc 3060
Db 3001 gatcaaatggcttacttgctgggttaattggctcactgtatcaccacaaaacccacctaacc 3060
QY 3061 aaagacaagtttaaaagtggtggaattgaaattaccaaacactggcattaacagcaggtggt 3120
Db 3061 aaagacaagtttaaaagtggtggaattgaaattaccaaacactggcattaacagcaggtggt 3120
QY 3121 aaaaagattaccaacattcaatcaggtgatattaccacaaaacagcaaatgatgctgtgaca 3180
Db 3121 aaaaagattaccaacattcaatcaggtgatattaccacaaaacagcaaatgatgctgtgaca 3180
QY 3181 ggcggtcggttttatgatttaaaacccgaaacttgaaagcaaaaatcaacagtgctgctaaa 3240
Db 3181 ggcggtcggttttatgatttaaaacccgaaacttgaaagcaaaaatcaacagtgctgctaaa 3240
QY 3241 acagcacaaaacattacacgaattctcagtagcagatgaacaaaggttaatcacatttacg 3300
Db 3241 acagcacaaaacattacacgaattctcagtagcagatgaacaaaggttaatcacatttacg 3300
QY 3301 gttagtaaccttactccagttatgacacctcaaaagacctctgatgctcacacctttgca 3360
Db 3301 gttagtaaccttactccagttatgacacctcaaaagacctctgatgctcacacctttgca 3360
QY 3361 ggtgaaaacggcattaccacaaagggttaataaagggttggtgctggtggcgattgaccaa 3420
Db 3361 ggtgaaaacggcattaccacaaagggttaataaagggttggtgctggtggcgattgaccaa 3420
QY 3421 accaaaagcttaaccacgacctaaagctgacggtgggttaataataatgaagcaggtatgctc 3480
Db 3421 accaaaagcttaaccacgacctaaagctgacggtgggttaataataatgaagcaggtatgctc 3480
QY 3481 attgacagtaaaagatggtcaaaaataccatcacaggaactaaagcaacacctctagctaaagt 3540
Db 3481 attgacagtaaaagatggtcaaaaataccatcacaggaactaaagcaacacctctagctaaagt 3540
QY 3541 accaatgattggtcagagacacgacctaaagcgaagggcttgccaatgacacgcgacaaaacc 3600
Db 3541 accaatgattggtcagagacacgacctaaagcgaagggcttgccaatgacacgcgacaaaacc 3600
QY 3601 cgtgcccagcagcttggtgatgcttaaaacgacgctttaacttgcaagggcaatgggtgaa 3660
Db 3601 cgtgcccagcagcttggtgatgcttaaaacgacgctttaacttgcaagggcaatgggtgaa 3660
QY 3661 gcggttgacttctcctcacttatgacacttgacacttgactttatcgatggcaatggcacacc 3720
Db 3661 gcggttgacttctcctcacttatgacacttgacacttgactttatcgatggcaatggcacacc 3720
QY 3721 gctaagtgacctatgatgacacaaagcaaaacccagtaaaagtggcttatgctgtaagt 3780
Db 3721 gctaagtgacctatgatgacacaaagcaaaacccagtaaaagtggcttatgctgtaagt 3780
QY 3781 gataataaaacattgaagtgaacagtgataaaaaacttggcgtcaaaaacccacacactg 3840
Db 3781 gataataaaacattgaagtgaacagtgataaaaaacttggcgtcaaaaacccacacactg 3840
QY 3841 accaaaacaaagtgttaattggttaacacaaaatttagtgcgcgcatggcgatggccctt 3900
Db 3841 accaaaacaaagtgttaattggttaacacaaaatttagtgcgcgcatggcgatggccctt 3900
QY 3901 gttaaagccagtgatgacgaccccatctaaataaccttggctggcgacatcccaaacgcgc 3960
Db 3901 gttaaagccagtgatgacgaccccatctaaataaccttggctggcgacatcccaaacgcgc 3960
QY 3961 aaaggggcaagcgaagcaagcagctcagcagctatgtggatgctgtagcacaagaagtc 4020
Db ||||| aaaggggcaagcgaagcaagcagctcagcagctatgtggatgctgtagcacaagaagtc 4020

QY 4021 atctatgacagtaccgataaagaagtactatcatcaagtcaatgacaagggtcaagtggacaaa 4080
|||||
Db 4021 atctatgacagtaccgataaagaagtactatcatcaagtcaatgacaagggtcaagtggacaaa 4080
QY 4081 acaaaagaagtggccaaagacaaaactgctgcacaaagcccaaaacccagatggcacattg 4140
|||||
Db 4081 acaaaagaagtggccaaagacaaaactgctgcacaaagcccaaaacccagatggcacattg 4140
QY 4141 gctcaaaatgaattgcaaaatcagctcatcaacaaagagcaagttaaatgatgccaataaaaag 4200
|||||
Db 4141 gctcaaaatgaattgcaaaatcagctcatcaacaaagagcaagttaaatgatgccaataaaaag 4200
QY 4201 caagatcaatgaagacacacccctttatcaaaagggcttgaaaacgcgcgaagacacc 4260
|||||
Db 4201 caagatcaatgaagacacacccctttatcaaaagggcttgaaaacgcgcgaagacacc 4260
QY 4261 aaaaacaaaacgcgcgacgttaactggtgatttaaatgcccgttgcccaaacacgcgtg 4320
|||||
Db 4261 aaaaacaaaacgcgcgacgttaactggtgatttaaatgcccgttgcccaaacacgcgtg 4320
QY 4321 acctttgcaggggatcacaggcaacgcgctaaaaactggggcgagactttgacctcaaa 4380
|||||
Db 4321 acctttgcaggggatcacaggcaacgcgctaaaaactggggcgagactttgacctcaaa 4380
QY 4381 ggtgggcaaacagacacccaataagctaaagcgcataataacatcggtggtgtagcaggtact 4440
|||||
Db 4381 ggtgggcaaacagacacccaataagctaaagcgcataataacatcggtggtgtagcaggtact 4440
QY 4441 gatggcttcaactgcaaaactggccaaagacacctaaccatctttaacagcgttaatgcaggt 4500
|||||
Db 4441 gatggcttcaactgcaaaactggccaaagacacctaaccatctttaacagcgttaatgcaggt 4500
QY 4501 ggcacagaattgatgaaaaagcatctctttgttagacgcaaacggtcaagccaaaagca 4560
|||||
Db 4501 ggcacagaattgatgaaaaagcatctctttgttagacgcaaacggtcaagccaaaagca 4560
QY 4561 aacacccctgcttaagtgcgaatggcctggacactgggtggcaaacgcgcacgttaacatc 4620
|||||
Db 4561 aacacccctgcttaagtgcgaatggcctggacactgggtggcaaacgcgcacgttaacatc 4620
QY 4621 ggtgcagctgtgatgaatacagatcggtgaactttaagcagcttaataagttgccaag 4680
|||||
Db 4621 ggtgcagctgtgatgaatacagatcggtgaactttaagcagcttaataagttgccaag 4680
QY 4681 acggtcaacaccttaaacacaaaagtaactcaggtgcgtcattacccctttgtgtaacc 4740
|||||
Db 4681 acggtcaacaccttaaacacaaaagtaactcaggtgcgtcattacccctttgtgtaacc 4740
QY 4741 gatgccaatggcaagcccaataaatggcaccgagtggaagcccaaaaagccatcaagggc 4800
|||||
Db 4741 gatgccaatggcaagcccaataaatggcaccgagtggaagcccaaaaagccatcaagggc 4800
QY 4801 gccgatgtaataactatcacgcaacgcgaacggcgctacctgtggacaagaatggcaag 4860
|||||
Db 4801 gccgatgtaataactatcacgcaacgcgaacggcgctacctgtggacaagaatggcaag 4860
QY 4861 cccatcacccgatgggacaaaacttgccaaatctggcagctcatggcaaaccccttgatgca 4920
|||||
Db 4861 cccatcacccgatgggacaaaacttgccaaatctggcagctcatggcaaaccccttgatgca 4920
QY 4921 ggtcatcaagtgggtggaagcctaggcggaactcagatgccatcacccctaaccacatc 4980
|||||
Db 4921 ggtcatcaagtgggtggaagcctaggcggaactcagatgccatcacccctaaccacatc 4980
QY 4981 aagtcacactttgccaacaaatggacacacacacacacaggttaatgccaatgagggcaagcc 5040
|||||
Db 4981 aagtcacactttgccaacaaatggacacacacacacacaggttaatgccaatgagggcaagcc 5040
QY 5041 caaagtctgccagcctcatcagcagcagcaagcaaaagttaatgctgcagtgtaaaagatgtg 5100
|||||
Db 5041 caaagtctgccagcctcatcagcagcagcaagcaaaagttaatgctgcagtgtaaaagatgtg 5100

QY 5101 ctaaaatgtaggcttttaacttgcagacccaatacctaagtggactttgttcaaaacctat 5160
|||||
Db 5101 ctaaaatgtaggcttttaacttgcagacccaatacctaagtggactttgttcaaaacctat 5160
QY 5161 gatccgctcaactttgtcaatggttacaggtgcgacatcacaaagcgtgctgtagtctgat 5220
|||||
Db 5161 gatccgctcaactttgtcaatggttacaggtgcgacatcacaaagcgtgctgtagtctgat 5220
QY 5221 ggcacgatgagtaacatcacctgcaacacgccttgacgagcagccgatgatgagcaat 5280
|||||
Db 5221 ggcacgatgagtaacatcacctgcaacacgccttgacgagcagccgatgatgagcaat 5280
QY 5281 gttcttataaaagccaaagatggttaagttctcaaaagcagacacatcgtccaaacgcgc 5340
|||||
Db 5281 gttcttataaaagccaaagatggttaagttctcaaaagcagacacatcgtccaaacgcgc 5340
QY 5341 tcactaaaacagcagcaaatcagccagtgatgccaactccaaactggtcttaagccttgtt 5400
|||||
Db 5341 tcactaaaagcagcgaatacagccagtgatgccaactccaaactggtcttaagccttgtt 5400
QY 5401 aaccccaatgctggttaaaagcagtcacaggcgatgagtgagtggtctcttaataacttatcaaaa 5460
|||||
Db 5401 aaccccaatgctggttaaaagcagtcacaggcgatgagtgagtggtctcttaataacttatcaaaa 5460
QY 5461 gcggtatttaataccaagatggtacaaactactaccacagtaagctctgagtgcatcagt 5520
|||||
Db 5461 gcggtatttaataccaagatggtacaaactactaccacagtaagctctgagtgcatcagt 5520
QY 5521 atccaaagcgaagataacagcagcatcaccttaagcgaatgggctgtaggcgtt 5580
|||||
Db 5521 atccaaagcgaagataacagcagcatcaccttaagcgaatgggctgtaggcgtt 5580
QY 5581 aaggtcatcaacaaatgtgggttaaaagcacaacaaagacacacgcgtgcgaatgtacaacag 5640
|||||
Db 5581 aaggtcatcaacaaatgtgggttaaaagcacaacaaagacacacgcgtgcgaatgtacaacag 5640
QY 5641 ttaaacgaagtacgcaacttgttggtctgtgtaagtctggttaataatgataacgctgacggc 5700
|||||
Db 5641 ttaaacgaagtacgcaacttgttggtctgtgtaagtctggttaataatgataacgctgacggc 5700
QY 5701 aatcagggtaaacattgcccagacatcaaaaaagacccaaattcaggttcatcttaaccgc 5760
|||||
Db 5701 aatcagggtaaacattgcccagacatcaaaaaagacccaaattcaggttcatcttaaccgc 5760
QY 5761 actgtcatcaaacgacgacacggttacttggcgtaaggttaaggttaataacgataccgaaaaactt 5820
|||||
Db 5761 actgtcatcaaacgacgacacggttacttggcgtaaggttaaggttaataacgataccgaaaaactt 5820
QY 5821 gccactggtggtgtgtacaaagtggcggtgataaaagcgcgaacgctaaacggcgttaagc 5880
|||||
Db 5821 gccactggtggtgtgtacaaagtggcggtgataaaagcgcgaacgctaaacggcgttaagc 5880
QY 5881 aatggttgggtcaaaaccccaaaaagatggcagcaaaaagccctgctgcacacttataac 5940
|||||
Db 5881 aatggttgggtcaaaaccccaaaaagatggcagcaaaaagccctgctgcacacttataac 5940
QY 5941 gccgaggttcagacaaactatttgaccaacaccccgagagaagccattgacagataaat 6000
|||||
Db 5941 gccgaggttcagacaaactatttgaccaacaccccgagagaagccattgacagataaat 6000
QY 6001 gaacaagggtatccgctcttcttcacgttcaacgattgcaaatcaagagcctgtgtgtaacagg 6060
|||||
Db 6001 gaacaagggtatccgctcttcttcacgttcaacgattgcaaatcaagagcctgtgtgtaacagg 6060
QY 6061 cgtaacggcatctgactcaagtgcctcaggaagcactcagtgcgatagagtttccaggcc 6120
|||||
Db 6061 cgtaacggcatctgactcaagtgcctcaggaagcactcagtgcgatagagtttccaggcc 6120
QY 6121 aaggcagatggtgaaagccgcgttgccatagggcagacaaaccccaagcaggaaccaatcc 6180
|||||
Db 6121 aaggcagatggtgaaagccgcgttgccatagggcagacaaaccccaagcaggaaccaatcc 6180
QY 6181 atcgccatcggtgtataacgcacaagccacgggcyatcaatcccatcgccatcggtacaggc 6240


```

Db      5992  cctggcagatccatcgggtaccgggggtatggccaccacacacggtcgaaggctgcggtggcca 6051
      |||
Qy      6835  gtgggactgtcgaagctgcgataatgggtcaatgggtattttaaatacaatgggttcagcc 6894
      |||
Db      6052  gtgggactgtcgaagctgcgataatgggtcaatgggtattttaaatacaatgggttcagcc 6111
      |||
Qy      6895  gatacccaagcccatgtagggcgcgagtggtgcaggttttcacttt 6942
      |||
Db      6112  gatacccaagcccatgtagggcgcgagtggtgcaggttttcacttt 6159
      |||

RESULT      3
AAF28545
ID      AAF28545 standard; DNA; 62909 BP.
XX
XX      AAF28545;
XX
XX      04-APR-2001 (first entry)
XX
XX      Genomic fragment #32.
XX
XX      Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW      bronchopulmonary; endocarditis; meningitis; ss.
XX
XX      Moraxella catarrhalis.
OS
XX      WO200078968-A2.
XX
XX      28-DEC-2000.
XX
XX      16-JUN-2000; 2000WO-US16649.
XX
XX      18-JUN-1999; 99US-0140121.
XX
XX      (INCY-) INCYTE GENOMICS INC.
XX
XX      Lagace RE, Patterson C, Berg KL;
PI
XX      WPI; 2001-041427/05.
XX
XX      Genomic library for identifying diagnostic and therapeutic
PT      compositions, and for identifying virulence factors, regulatory
PT      elements and drug targets, comprises Moraxella catarrhalis nucleic
PT      acids -
XX
XX      Claim 1; Page 293-308; 545pp; English.
XX
XX      The present invention relates to a Moraxella catarrhalis genomic library
CC      comprising a combination of 41 nucleic acid molecules (see
CC      AAF28514-AAF28554). The library has a number of uses described in the
CC      specification e.g. is useful for identifying diagnostic and therapeutic
CC      compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC      aerobic, gram-negative diplococcus, normally found among the bacterial
CC      flora of human upper airways. M. catarrhalis is known to cause acute,
CC      localised infections such as otitis media, sinusitis and bronchopulmonary
CC      infection and life-threatening, systemic diseases including endocarditis
CC      and meningitis.
XX
XX      Sequence 62909 BP; 18449 A; 13798 C; 12744 G; 17918 T; 0 other;
SQ

Query Match      41.1%; Score 2852.4; DB 22; Length 62909;
Best Local Similarity 69.3%; Pred. No. 0;
Matches 4853; Conservative 0; Mismatches 991; Indels 1156; Gaps 25;

Qy      1 atgaatcacatctataaagtcatctttaaacaagccacagggcacattatggccgtggcca 60
      |||
Db      42630 atgaatcacatctataaagtcatctttaaacaagccacagggcacattatggccgtggcg 42689
      |||
Qy      61 gagtgcgcccaatccccacagc-ggaggggatagcagtagtaccgcaggacaggtgggcag 119
      |||
Db      42690 gaatatgcgaatccccacagcagcgggggggatagctgtgtctacagggcaagtggcag 42749
      |||

```


Db 4144 gcaagccaagcgaacaactcagcaggcgtatgtgagctgatggccaataaagggtcatctat 4203
Qy 4027 gacagtaccgataagaagtactatcaagtcaatgacaagggtgcaagtgggacaaaacaaa 4086
Db 4204 gacagtaccgataagaagtactatcaagcgaacaaatgatggcacagttgataaaaaccaa 4263
Qy 4087 gaagtccaaaagacaactggctgcgccaagcccaaaccccccagatggcacatlggctcaa 4146
Db 4264 gaagtccaaaagacaactggctgcgccaagcccaaaccccccagatggcacatlggctcaa 4323
Qy 4147 atgaatgcaaatcagtcattacaagaagcgaagtaaatgatgccaataaaaagcaaggc 4206
Db 4324 atgaatgcaaatcagtcattacaagaagcgaagtaaatgatgccaataaaaagcaaggc 4383
Qy 4207 atcaatgaagacacgacctttatcaaaaggcttgaacccgcccgaagacacacaaacc 4266
Db 4384 atcaatgaagacacgacctttttaaaggacttgaagaagccgcttctgatacaaaaacc 4443
Qy 4267 aaaaacccgagtaactgtgggtgatttaaatgccgttcccaaaaacacgcgtgaccttt 4326
Db 4444 aaaaacccgagtaactgtgggtgatttaaatgccgttcccaaaaacacgcgtgaccttt 4503
Qy 4327 gcagggtatcagcgaacacggtcaaaaactggcgagactttgaccatcaaaaggtggg 4386
Db 4504 gcaagggatcacaggcaacacggtcaaaaactggcgagactttgaccatcaaaaggtggg 4563
Qy 4387 caaacagacacccaataagctaacgcgataataacatcggtgtgtagcaggtactgtggc 4446
Db 4564 caaacagacacccaataagctaacgcgataataacatcggtgtgtagcaggtactgtggc 4623
Qy 4447 ttaactgcaaaacttgccaagaacctaaccaatttaacagcgttaatgacaggtggcacc 4506
Db 4624 ttaactgcaaaacttgccaagaacctaaccaatttaacagcgttaatgacaggtggcacc 4683
Qy 4507 agaattgataaaaagcatctctttgttagcgaacacggtcaagcgaacaaacacc 4566
Db 4684 aaaattgatcaaaagcgtgtcttttgtagactcaagcgttcaagcgaacaaacacc 4743
Qy 4567 cctgtctaagtgcgaatggctggacacctgggtggcgaacgcgcatcagtaacatcggtgca 4626
Db 4744 cctgtctaagtgcgaatggctggacacctgggtggc----- 4779
Qy 4627 gctgttatgataaacgtagcgggtgaactttaagcagtttaatgaagtgcgaacacggtc 4686
Db 4780 ----- 4779
Qy 4687 aacaactaaacacaaagtaactaactgagtgcgctattaccctttgtgtaacgcgagtc 4746
Db 4780 ----- 4779
Qy 4747 aatggcaagcccatcaatggcacccgatggcaagcccaaaaagccatcaaggcgccgat 4806
Db 4780 ----- 4779
Qy 4807 ggtataactatcacgccaacgcaacgcgtaacctgtgacaaaagatggcaagcccatc 4866
Db 4780 ----- 4779
Qy 4867 accgatcggaacaaacttgccaaactgtggcgctcatgtggcaaaccccttgatgacggtcat 4926
Db 4780 ----- 4779
Qy 4927 caagtggtggaagccttagcggaactcagatgccaatcaccctcaacacatacaagtc 4986
Db 4780 ----- 4779
Qy 4987 actttgcacaaattgacacacaaacacagggtaagtccaatgacagggcaagcccaagt 5046
Db 4780 ----- 4779
Qy 5047 ctgcccagcctatcagcagcagcaagtaagtgtgctcaggtgtcaaaagtgtgctaaat 5106
Db 4780 ----- 4779

Qy 5107 gtaggctttaacttgcagaccaatcacaaatcaagtggactttgtcaaaagcctatgatacc 5166
Db 4780 ----- 4779
Qy 5167 gtcaactttgtcaatggtacaggtgcccagacatcacaaagcgtgcgtagtgcgtgacg 5226
Db 4780 ----- 4779
Qy 5227 atgagtaacatcacccgtcaaacaccgccttagcagcgaccgatgatggcgaatgtgctt 5286
Db 4780 ----- 4779
Qy 5287 atcaaaagcgaagatggtlaagtctacaagcagcagacctcatgccaacggtcacta 5346
Db 4780 ----- 4779
Qy 5347 aaagcagggcaaatcagccagtgatgccaacaaactccaactggctaaagccttgttaacccc 5406
Db 4780 ----- 4779
Qy 5407 aatgctggtaaaggcagtcacagcgatgcgtggctcttaataacttatcaaaagcggt 5466
Db 4780 ----- 4779
Qy 5467 tttaataccaaagatggtacaacttaccacagtaagctctgtaggcacagtatccaa 5526
Db 4780 ----- 4779
Qy 5527 ggcgaagataacagcagcatcaccctaaagcaaatgggtgaatgtaggcggtaaggtc 5586
Db 4780 ----- 4785
Qy 5587 atcagcaatgtggtaaaagcacaacaaagacaccgacgctgccaatgtacaacagttaaac 5646
Db 4786 atcagtaatgtggcaaaagcacaacaaagataccgacgctgccaatgtacaacagttaaac 4845
Qy 5647 gaagtacgcgaacttgggtcttggtaagtgtgtaacgctgataaacgctgacggcaatcag 5706
Db 4846 gaagtacgcgaacttgggtcttggtaagtgtgtaacgctgataaacgctgacggcaatcag 4905
Qy 5707 gtaaacattgcccagatcaaaaaagaccccaattcaggttcatctataacccgacgtc 5766
Db 4906 gtaaacattgcccagatcaaaaaagaccccaattcaggttcatctataacccgacgtc 4965
Qy 5767 atcaaacgagcagcgttacttggcgtaaaagtaataacgataccgaaaacttgcacact 5826
Db 4966 atcaaacgagcagcgttacttggcggtaaaggtataacgataccgaaaacttgcacact 5025
Qy 5827 ggtgggtgtacaagtggcggtggataaaagcggcaacgctaacgcgatttaagcaatgtt 5886
Db 5026 ggtgggtgtacaagtggcggtggataaaagcggcaacgctaacgcgatttaagcaatgtt 5085
Qy 5887 tgggtcaaaaaccccaaaaagatggcagcaaaaacccctgctgcacattataacgcgcga 5946
Db 5086 tgggtcaaaaaccccaaaaagatggcagcaaaaacccctgctgcacattataacgcgcga 5145
Qy 5947 ggtcagacccaactatttgaccaacaaccccgacagacccattgacagataaataaagaaca 6006
Db 5146 ggtcagacccaactatttgaccaacaaccccgacagacccattgacagataaataaagaaca 5205
Qy 6007 ggtatccgcttcttccatgtcaacgatggcaatcaaaagcctgtggtaacagggcgtaac 6066
Db 5206 ggtatccgcttcttccatgtcaacgatggcaatcaaaagcctgtggtaacagggcgtaac 5265
Qy 6067 ggcattgactcaaatgctcaggaacacactcagtgccgataagtttccagggcgaagga 6126
Db 5266 ggcattgactcaaatgctcaggaacacactcagtgccgataagtttccagggcgaagga 5325
Qy 6127 gatggtgaagcgcgcttgccatagggcagacaaaccccaagcaggaacccatccatgc 6186
Db 5326 gatggtgaagcgcgcttgccatagggcagacaaaccccaagcaggaacccatccatgc 5385

Dbb 1019 catttgcttgatcagcatgtaactcctaataacatccgaaaggttactctgtgattaacgatcttt 1078
Qy 523 attcaaaaccatcagattaccgccaaatacagagactcaaatgggtctc---gaaatat 579
Dbb 1079 attaaaggccatcgattataaagaataacgaagctcaaaagataatgatgataaataat 1138
Qy 580 agacgcacagcagcagaagacacgcccagctactgcagtgaggagccatgcatatgcaaat 639
Dbb 1139 agacgcacaacccgaagcagcagccagctactgcagtgaggagccatgcatatgcaag 1198
Qy 640 ggtcattttccaaacgctttgttacacggtcaacagctgaagcgaactattccttgcca 699
Dbb 1199 ggtcattttccaaacgctttgttacacggtcaacagctgaagcgaactattccttgcca 1258
Qy 700 gtagggttaccgccaaagccgaaagagatatacaatcgtattggtttcattatgcaaa 759
Dbb 1259 gtgggtcttcgcccacagcagcgaggccaaatctacaatcgtattggtttcattatgcaaa 1318
Qy 760 gctatcaattatgagcacactagcccttggtgcagatactcgtagttggttgcaggt 819
Dbb 1319 tctagctcgttgagagcga tagcccttggcgaggtactcgtcagctacagggcag 1378
Qy 820 attgccttaggttatggttctcagatccttaataataataataata---ataataaa 876
Dbb 1379 attgccttaggttaaggttctgttgcactcagagtgatataataattctagacggcctat 1438
Qy 877 gctatgtaccagaaggtaatgggtcaaacataaaaatcgtctaaagccacggcaatgg 936
Dbb 1439 acacaaatacccaggcactagaccaccaagtctcaagccaccaataacgaagcggt 1498
Qy 937 ttattttccattgtagcactatacaagcgrtaaaatcattcgtcgtgcaggttat 996
Dbb 1499 ccactttccattgtagtactctataaacgttaaaatcattcgtcgtgcaggttat 1558
Qy 997 gaggataccgctcgggtccaatgtggcacagctaaaaagcgtggagaaatcgtgcta---ag 1053
Dbb 1559 aataaaaccgctcgggtccaatgtggcacagctagaagcgggtggtaagtgggttaaggag 1618
Qy 1054 cgtcaaatattctttaaagggtgatgataacggttactggcgttaagaataaaactgggcgag 1113
Dbb 1619 cgtagaattactttcaggggtgataacagctactgacgtactgacgtactgacgtactgacgt 1678
Qy 1114 actttaaccattaaaggtgtagacccaagcggacaaagcgaactaacgtataataataaact 1173
Dbb 1679 actttaactataaagggtg---cagagaccaacgacttaaccgataataataatcgt 1735
Qy 1174 ggttggttaacagataataactggtcgtgaaagttaaaacttgcttaaaacacctaagcgt 1233
Dbb 1736 gtggtaaaaggcgtgataatagtggtcgtgaaagttaaaacttgcttaaaacacctaatacaat 1795
Qy 1234 ctgaaacaggttagcaccacaaacctaaccgcccagcgagaaagttagcgttag---gtagt 1290
Dbb 1796 cttagtgagggtgatacaactacataataatgagtgataagtttaacacacacaggttagttagt 1855
Qy 1291 ggttaataaccgctgagctacaaagcgttggttttaacacacacacacacacacacacacac 1344
Dbb 1856 agtagtactacagctgaattatgagtgatgagtgatgagtgagtgagtgagtgagtgagtgag 1915
Qy 1345 aatgcagcagcagacaaacccgtctatggtcagctgagtgagtgagtgagtgagtgagtgagtg 1404
Dbb 1916 agtcaaaagcagacaaacccgtctatggttgagtgagtgagtgagtgagtgagtgagtgagtg 1975
Qy 1405 aatacggcacttga---gatactactcgtatcaccaaaagataaaaattggttttagcaat 1461
Dbb 1976 gaaacaacagcagaactcggcactactcgtattaccagagataaaaattggttttagcga 2035
Qy 1462 aagcgtgtagttagtgaacaaac 1516
Dbb 2036 gatgggtgtagttagtgaacaaac 2095
Qy 1517 ----- 1518
Dbb 2096 agtgttgcaattaccatagacaatggcattgtagcaggttaataaaagatcagtaactctt 2155

Qy 1519 ggaacagcaccctaaacaaagcgttggtgactgtttaataacaccattggtgggtagcaat 1578
Dbb 2156 gccaaaggtagcagtgtaaacagtgcggttaccatcgacagctcaaaagccgccaagcct 2215
Qy 1579 aaacaaatccaagtcggtgctgagtgcatataaatttcccgatgtgaatgttaattatca 1638
Dbb 2216 acttaaacagcagcgtgcgcagtgctcacacctactgaaatcatcagttgagtgtaag 2275
Qy 1639 aatgcgcacaaatcgcgactactcgtattacggaagagaaattggttttgcgtgagct 1698
Dbb 2276 agtggcaatgttaccgcccacacttacaacattggcgtgaaacaccagcgttaacagt 2335
Qy 1699 gatggttaagtgtgataa----- 1715
Dbb 2336 gatggcactagtataaatttagtgttaagggttagtggtacgaaacatagcttagttacc 2395
Qy 1716 -----aaagtcaccattttggataaaaaaacaacttcaagtggtgggtggtt 1761
Dbb 2396 gccgaacatttggcagctactataatggaagtcacatgaaacggtgcagctgctctcaa 2455
Qy 1762 aaattaccacaaagacagtgccattaatgcaggtgatcaaaagatcagtaattgttaaaagt 1821
Dbb 2456 agctttaccgttaaaagaagacagatgagcgcacacgtatcacgcgtggtggttaagat 2515
Qy 1822 gcaacgacatgacgtacgtacgtacttaataacagcttaacacaggtcccaacagacgcc 1881
Dbb 2516 acgacaaaaaatgcggcgagtcagcatcttaaaactcaaaaggtataaaacgggtctaagc 2575
Qy 1882 gacggtgcccataaa-----agcttctctattcgtgataaaaaaggttcagga 1929
Dbb 2576 gttgctaccacaaagatggttacggttacccttgggttagcgaagtagcgtgctgacc 2635
Qy 1930 ttacgattagtaactgtattctaatggttaataaccacaaatcctttgagaccatcacc 1989
Dbb 2636 atggcaaaagcaccctaaacacagatgggtgactgttaaaagataccaaacgaacaaatc 2695
Qy 1990 ttgtaggtgaaacggcgtacgtatcagcaatgacatagcgaaggttaaaaggttaaaag-- 2047
Dbb 2696 caagtcggtgtaagtgcattaaatttactaatgtaagttagtaataccaggtactggc 2755
Qy 2048 -ttggtattgacccaatcaatggtctcacacgctgaagcgtgacggtgaggtgagcgaataa 2106
Dbb 2756 atggcaataccgctcgtcattaccagagataaaaattggttctggttctgagtggtgca 2815
Qy 2107 gatgtaaaaactcaattggtttattgagcaag----- 2137
Dbb 2816 gttgatacaacaaaccttattcttgatcaagacaagctacaagttggcaattgtaagatt 2875
Qy 2138 ----tggtgagcgttaacgac 2193
Dbb 2876 accaaactggcattaaacgaggtggtaagccatcacagggtgtccccaacacactgct 2935
Qy 2194 agcattaccatgaggtggtgctacgacacacacacacacacacacacacacacacacacacac 2253
Dbb 2936 agcattccgatacaagtagc---cgcaacatagaactgggcaatacaatccca---agac 2989
Qy 2254 gaagacaaatccaaagcgcgagctatcgtgataataataacacagcgttttaacctaaaa 2313
Dbb 2990 aaagacaaatccaaacgctgccagcattaaatgataataataacacagcgttttaacctaaaa 3049
Qy 2314 aataatgacacgtggtggttctgctccacttataacacactgttgacttttctatgagtcgc 2373
Dbb 3050 aataataac 3109
Qy 2374 aatgcac 2433
Dbb 3110 aatgcac 3169
Qy 2434 gatgtcaatgtgagtaaaac 2493
Dbb 3170 gatgtgagtggtgagatacaaac 3226

Db 5321 ----- 5320
QY 4687 acaacctaacaacaaagtaactacgtgctgctattaccctttgtggtaacccgatgcc 4746
Db 5321 ----- 5320
QY 4747 aatggcaagcccatcaatggaacccgatggcaagcccaaaaaagccatcaagggcgccgat 4806
Db 5321 ----- 5320
QY 4807 ggtataactatcacgccaacgccaacgcttacctgttgacaaaagatggcaagcccatc 4866
Db 5321 ----- 5320
QY 4867 accgatgcgacaaaacttgccaatctggcagctcatggcaaaccccttgatgcaggtcat 4926
Db 5321 ----- 5320
QY 4927 caagtggcgcaagccttagggcggaactcagatgccalcacctaacccaacatcaagttcc 4986
Db 5321 ----- 5320
QY 4987 actttgccaaaattgacacacacacacaggttaatgccaatgcaggggcaagcccaagt 5046
Db 5321 ----- 5320
QY 5047 ctgcccagcctatcagcagcacagcaaaagttaatgctgccagtgccaagatgtgctaaat 5106
Db 5321 ----- 5320
QY 5107 gtaggctttaactgacagaccaatcacatcaagtggactttgtcaaaagcctatgatacc 5166
Db 5321 ----- 5320
QY 5167 gtcaactttgtaagtgtacaggtgcgcacatcacaaagctgcgtagtgctgatggcaag 5226
Db 5321 ----- 5320
QY 5227 atgagtaacatcacctgccaacaccccttagcagcgaccgatgatgatggcaatgtgctt 5286
Db 5321 ----- 5320
QY 5287 atcaaaagcaaaagatggaagtctctacaaagcagacacatcagcaaacgcgctacta 5346
Db 5321 ----- 5320
QY 5347 aaagcaggcaaatcagccagtgatgccaaaactccaactggttctaagccttgttaacccc 5406
Db 5321 ----- 5320
QY 5407 aatgctggtaaagcagtaacagcgatgcagtgagtgcttctaataacttatcaaaagcgta 5466
Db 5321 ----- 5320
QY 5467 tttaaatccaaagatggtacaaactactaacacagtaagctctgatggcatcagtatccaa 5526
Db 5321 ----- 5320
QY 5527 ggcaaaagataacagcagcatcacccctaaagcaaatggctgaatgtaggcggttaaggtc 5586
Db 5321 ----- 5326
QY 5587 atcagcaatggtggttaagacacaaaagacacccgacgctgccaatgtacacagttaaac 5646
Db 5327 atcagtaatgggcaaaagcacaagataccgacgctgccaatgtacacagttaaac 5386
QY 5647 gaagtaacgaactgttggtgtcttggttaatgctggtaatgatcaacgctgcagcgaatcag 5706
Db 5387 gaagtacgcaactgttggtgtcttggttaatgctggtaatgatcaacgctgcagcgaatcag 5446
QY 5707 gtaaacattgccacatcaaaaagacccaaattcaggttcatcatcattacccgactgtc 5766
Db 5447 gtaaacattgccgacatcaaaaagacccaaattcaggttcatcatcattacccgactgtc 5506

QY 5767 atcaaaagcagcggtacttgctggtggttaaaggtaataacgataccgaaaaacttgcact 5826
Db 5507 atcaaaagcagcggtacttgctggtggttaaaggtaataacgataccgaaaaacttgcact 5566
QY 5827 ggtggtgtacaagtggcggtggataaaagcggcaacgctaacgcttaacggcgatattaaagcaattgtt 5886
Db 5567 ggtggtatacaagtggcggtggataaaagcggcaacgctaacgcttaacggcgatattaaagcaattgtt 5626
QY 5887 tgggtcaaaacccaaagatggcagcaaaaagccctgctgcacatttaaacgcgcga 5946
Db 5627 tgggtcaaaacccaaagatggcagcaaaaagccctgctgcacatttaaacgcgcga 5686
QY 5947 ggtcagaccactatttgaccaacaaccccgagaaagccattgacagaataaaatgaacaa 6006
Db 5687 ggtcagaccactatttgaccaacaaccccgagaaagccattgacagaataaaatgaacaa 5746
QY 6007 ggtatccgcttcttccatgtcaacgatggcaatcaaaagcctgtggtacaagggcgtaac 6066
Db 5747 ggtatccgcttcttccatgtcaacgatggcaatcaaaagcctgtggtacaagggcgtaac 5806
QY 6067 ggcattgactcaagtgcctcagcaagcactcagtgccgataggtttccagggccaagga 6126
Db 5807 ggcattgactcaagtgcctcagcaagcactcagtgccgataggtttccagggccaagga 5866
QY 6127 gatggtgaagcccgcttgccatagggcagacaacccaagcaggaacccaatccatcgcc 6186
Db 5867 gatggtgaagcccgcttgccatagggcagacaacccaagcaggaacccaatccatcgcc 5926
QY 6187 atcgggtgataacgcacaagccagggcgatcaatccatcccatcggtacaagggcaattgtg 6246
Db 5927 atcgggtgataacgcacaagccagggcgatcaatccatcccatcggtacaagggcaattgtg 5986
QY 6247 gtaacaggtgaagcactctgctgcacgcgacccaagcactgttaaggtgtataacagt 6306
Db 5987 gtagcaggtgaagcactctgctgcacgcgacccaagcactgttaaggtgtataacagt 6046
QY 6307 tacagtggtggttaataacaacccagtttatcgatgccactcagacccgatgtctttggtgtg 6366
Db 6047 tacagtggtggttaataacaacccagtttatcgatgccactcagacccgatgtctttggtgtg 6106
QY 6367 ggaataacatcacctgacgcggaagtaactcggttccttaggttcaaacctcgcacatc 6426
Db 6107 ggaataacatcacctgacgcggaagtaactcggttccttaggttcaaacctcgcacatc 6166
QY 6427 agtgacggcacacgcaggcacacacagccaataatctgacggcacagcaggtacaaac 6486
Db 6167 agtgacggcacacgcaggcacacacagccaataatctgacggcacagcaggtacaaac 6226
QY 6487 accacagcaggtgcaacaggtacgggttaaggtttgtggaacaaacggcggttgggtgcg 6546
Db 6227 accacagcaggtgcaacaggtacgggttaaggtttgtggaacaaacggcggttgggtgcg 6286
QY 6547 gtcctcggtgggtgctcaggtgctgaacgcgcgtatccaaaatgtggcagcaggtggaagtc 6606
Db 6287 gtcctcggtgggtgctcaggtgctgaacgcgcgtatccaaaatgtggcagcaggtggaagtc 6346
QY 6607 agtgccaccagcagcaggtgctcaatggtgagcagttgtacaaaagccacccagggcatt 6666
Db 6347 agtgccaccagcagcaggtgctcaatggtgagcagttgtacaaaagccacccagggcatt 6406
QY 6667 gccaaacgaacccaatgagcttgaccatcgatccacccaaacgaataaaagcccaatgca 6726
Db 6407 gccaaacgaacccaatgagcttgaccatcgatccacccaaacgaataaaagcccaatgca 6466
QY 6727 gggatttcatcagcgatggcgatggcgatgccaaagcctcatctctgagcagatcc 6786
Db 6467 gggatttcatcagcgatggcgatggcgatgccaaagcctcatctctgagcagatcc 6526
QY 6787 atggttaccgggggtattgcccacccacaagcgtcaaggtggtggcggtgggactgtcg 6846
Db 6527 atggttaccgggggtattgcccacccacaagcgtcaaggtggtggcggtgggactgtcg 6586


```
QY 5346 aaagcaggcaaatcagccagtgatgcacaaactccaactggtctaagccttgtaaccc 5405
Db 5320 ----- 5319
QY 5406 caatgctggttaagggcagtagcagggcagtgagtggtctttaataacttatcaaaagcgt 5465
Db 5320 ----- 5319
QY 5466 atttaaatccaaagatggttacaactactaccacagtaagctctgatggcatcagatcca 5525
Db 5320 ----- 5319
QY 5526 aggcagaagataaacagcagcatcacccctaagcaaaagtgggctgaatgtagcggtaaggt 5585
Db 5320 -----aaggt 5324
QY 5586 catcagcaatgtggttaaaagccacaaagacacgcagcgtgccaatgtacaacagttaaa 5645
Db 5325 catcagtaagtgtggcgaagacaaagataccgacgcgtgccaatgtacaacagttaaa 5684
QY 5646 cgaagtagcgaactgttggtgctcttggttaattgctggttaataagataacgctgacggcaatca 5705
Db 5385 cgaagtagcgaactgttggtgctcttggttaattgctggttaataagataacgctgacggcaatca 5444
QY 5706 ggttaacattgcacatcaaaaaaagcccaaatctcaggttcacatctacaacgcactgt 5765
Db 5445 ggttaacattgcacatcaaaaaaagcccaaatctcaggttcacatctacaacgcactgt 5504
QY 5766 catcaagcagcagcagctactgtggttaaaagtaataacgataccgaaaaacttgcac 5825
Db 5505 catcaagcagcagcagctactgtggttaaaagtaataacgataccgaaaaacttgcac 5564
QY 5826 tgggtgtcaaatggcgtggataaagacggcaacgcgttaacggcgtatttaagcaatgt 5885
Db 5565 tgggtgtcaaatggcgtggataaagacggcaacgcgttaacggcgtatttaagcaatgt 5624
QY 5886 ttgggtcaaaacccaaagataggcagcaaaaaagccctcgtcgcacttataacgcgc 5945
Db 5625 ttgggtcaaaacccaaagataggcagcaaaaaagccctcgtcgcacttataacgcgc 5684
QY 5946 aggtcagaccaaactatttgaccacaaccccgagagacccattgacagaataaatagaaca 6005
Db 5685 aggtcagaccaaactatttgaccacaaccccgagagacccattgacagaataaatagaaca 5744
QY 6006 aggtatccgtcttccatgtccaacgatggcaatccaagagcctgtgtgtaacaggcgtaa 6065
Db 5745 aggtatccgtcttccatgtccaacgatggcaatccaagagcctgtgtgtaacaggcgtaa 5804
QY 6066 cggcattgactcaagtccctcaggaagcactcagtggtgagatagggttccagggccaagtc 6125
Db 5805 cggcattgactcaagtccctcaggaagcactcagtggtgagatagggttccagggccaagtc 5864
QY 6126 agatgtgtaagccgcgttgccatagcagacaaacccaagcaggaaccaatccatcgc 6185
Db 5865 agatgtgtaagccgcgttgccatagcagacaaacccaagcaggaaccaatccatcgc 5924
QY 6186 catcgtgataaacgacaaagccagcagatccaatccatccatcgccatcggtacaggaatgt 6245
Db 5925 catcgtgataaacgacaaagccagcagatccaatccatccatcggtacaggaatgt 5984
QY 6246 ggttaacaggtaagcactctggtgcacatcggcagcccaagcactgtttaaggctgataacag 6305
Db 5985 ggttagcaggttaagcactctggtgcacatcggcagcccaagcactgtttaaggctgataacag 6044
QY 6306 ttacagtgtggttaataacacacagtttatcgatggccactcagacagatctcttctgtgt 6365
Db 6045 ttacagtgtggttaataacacacagtttatcgatggccactcagacagatctcttctgtgt 6104
QY 6366 gggcaataaacatcacctgacccgaagtaactcgtgttccttaggttcaaaactctgcat 6425
Db 6105 gggcaataaacatcacctgacccgaagtaactcgtgttccttaggttcaaaactctgcat 6164
```

```
QY 6426 cagtgcaggccacacacgacagcgacacaaagccaaataatctgacggccacagcaggtacaac 6485
Db 6165 cagtgcaggccacacacgacagcgacacaaagccaaataatctgacggccacagcaggtacaac 6224
QY 6486 caccacagcaggtgcaacacaggtacggttaaaagccttgctggacaaacgagcgttgatgc 6545
Db 6225 caccacagcaggtgcaacacggtacggttaaaagccttgctggacaaacgagcgttgatgc 6284
QY 6546 ggtctccgtggtgctcctcaggtgctgaacccgcttatccaaatgtgacagcaggtgaggt 6605
Db 6285 ggtctccgtggtgctcctcaggtgctgaacccgcttatccaaatgtgacagcaggtgaggt 6344
QY 6606 cagtgcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6665
Db 6345 cagtgcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6404
QY 6666 tgcacacgcaaccaatgagcttgaccatcgatggcgtccatccacaaacgaaataaagccaatgc 6725
Db 6405 tgcacacgcaaccaatgagcttgaccatcgatggcgtccatccacaaacgaaataaagccaatgc 6464
QY 6726 aggtatttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6785
Db 6465 aggtatttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6524
QY 6786 catggttacccggggtattgccaccccaaacgggtcaagggtgcgtgagcagtgagcagtc 6845
Db 6525 catggttacccggggtattgccaccccaaacgggtcaagggtgcgtgagcagtcagcagtc 6584
QY 6846 gaagctcgcagataatggtcaatgggtatttaaatcaatgggttcagcagcagcagcagcagc 6905
Db 6585 gaagctcgcagataatggtcaatgggtatttaaatcaatgggttcagcagcagcagcagcagc 6644
QY 6906 ccatgtaggggcgagcagctggtgcaggttttcaactt 6942
Db 6645 ccatgtaggggcgagcagctggtgcaggttttcaactt 6681

RESULT 7
AAF59100
ID AAF59100 standard; DNA: 6973 BP.
XX
AC AAF59100;
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kda gene SEQ ID NO:1.
XX
KW Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kda outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX
DR WPI; 2001-159722/16.
XX
P-PSDB; AAB69133.
XX
PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
XX
PS Example 3; Fig 2A-W; 247pp; English.
XX
```


Qy	1716	-----aaagtcaccataatttggataaaaaaaccaactccaagtgggtggtgt	1760
Db	2396	cgcgaacatttgccagctatctaataagaagtcactcgaacgcgtcgacagttgctctaca	2455
Qy	1761	taaaattaccaaaagacagtggcatlaatlgcagggtgatcaaaagatcagtaaatgttaaaga	1820
Db	2456	aagctttaccogtttaaaagaagaagatgtagcagccaacgctatcacctgtgcttaaga	2515
Qy	1821	tgcaacggacgataccgatagcagctgacgtacctataaaacagcttaaacaaagtccaacaagacgc	1880
Db	2516	tagacaaaaatgcccgcgcagtcagcatottaaaactcaaaggtaaaaacggtctaac	2575
Qy	1881	cgacggtgcctacaa-----agcttctctatttgttgatgaaaaaaggccaaga	1928
Db	2576	ggttgctaccaaaaaagatggtaacgttacctttgggtgttagccaagatgacggtctgac	2635
Qy	1929	atttagattagtaacctgttatcttaattggttaatacccccaaatacctttgagaccatcac	1988
Db	2636	cattggcaaaagcaccttaaaacaacagatggcttgactgttaaaatatcccaacgaacaaat	2695
Qy	1989	otttgaggtgaaaaacygcatacagatcacgaatgacatgccaaaagttaaagtcaaaag-	2047
Db	2696	ccaagtcggtgctaaTgcgcatTaaatttactaaTgaatggttagtaatccaggtactcg	2755
Qy	2048	--ttgatttgacccaatcaatgggtctcaacgcctcaagctgacgtgggttagcgataa	2105
Db	2756	cattgcaaatcacgcctcgcatcacagagataaaaattggcttctggtctctgatggtgc	2815
Qy	2106	agatggTaaaactcaattggttatlgagcaag-----	2137
Db	2816	agttgatacaaaaacaccttatctgatacgaagacagctacaagttggcaatgttaagat	2875
Qy	2138	----tggctagcgttaacgacaccaaaaaacatcattagaggtattgcccaacactgcc	2192
Db	2876	taccaacttggtcatTaacycgaggtggTaaagccaactcacaggtctgtccccaacactgcc	2935
Qy	2193	tagcattaccaatgacgtggcgtacgacacacagacagggccaatacataatcacacgca	2252
Db	2936	tagcatTgcgcgTcaaagTgc---cgcaacatagaactggcgcaatacaatcca---	2989
Qy	2253	cgaagacaaatccaaagccgcagtatcggtgatatataaatacaggtctttaacctaaa	2312
Db	2990	caaaagacaaatccaaacgcTgccagcatTaaTgatatTaaatacaggtctttaacctaaa	3049
Qy	2313	aaataatagcaactccggttggotTTgtctccacttataaacactgttgactttatcgatgg	2372
Db	3050	aaataaTaaacccaattTgactttgtctccacttatgacattgttgactttggccaatgg	3109
Qy	2373	caatgcaccacacgcTaaagTaaacttagatgaacccaatcaaacaccagtaaaagTaaactta	2432
Db	3110	caaTgcaccacacgcTcaagTaaaccaTgataccgcTcaaaaacccgTaaagTggTata	3169
Qy	2433	TgattGcaattGtgatgagaaaaaccattGaactcacagcgcgataatTggcaagacaacaa	2492
Db	3170	TgatTgaatTggtatgatacaacctTcaTctaacaggcactgatgacaa---Taaaaa	3226
Qy	2493	aattggcgTcaaaaaaccacacactgacacaacaaatTgctaatTggtTaaagcaacc----	2547
Db	3227	actTggcgTcaaaacccacccaactGaacaaaacaaagTctaTgtTaatatagcaacactaa	3286
Qy	2548	----aaCTttagTaccacgcgataaagcTgcccTgtTaaCgccaagagacatCgcgcgaaa	2603
Db	3287	CtttaatTgtTaaCtctagTgaTgaagatTgcccTgtTaaCgccaagagacataCgcgcgaaa	3346
Qy	2604	TctaaacacccTgcaagaagaaattCcacaccacaaaagccacagacacccgccttaca	2663
Db	3347	TctaaacacccTgcaagaagaaatTcaacaccacaaaagccacagacacccgccttaca	3406
Qy	2664	aacctTtaagtCa aaaagacggtTgca-----actgatTgacgaaacccataccgTggg	2717
Db	3407	aacctTtacogttTaaaaaggtTagaTgaaaataaTaatTgctgatCgacgcaacgcatacc	3466

Qy	2718	taaaga tgggtacacaaaaacggaagccgtcaaacctctataaactcaaaagtgaacaacg	2777
Db	3467	cgttggtcaaaaaagacgcgaataatcaagctcacaccctaacctcaaaggctgaacaacg	3526
Qy	2778	tctaagcgttgcctaccaataaagatggttacgttacctttggcattaacacccaagaacg	2837
Db	3527	tcttaattataaacacgcacaaaatggttcggttacctttggcatlaacaccacaacg	3586
Qy	2838	tcttaaagccgcgcagcaccaactctaaacaaagtgcgttgtctattaataaaccccgc	2897
Db	3587	tcttaagccgcgcacaaagcacc--ctaaacgcggtggttgtctttaaaaaacccac	3643
Qy	2898	tagtaacgcacaaatccaagtcggtctgtgatggcgtagaagtttgccaaggt---tgataa	2954
Db	3644	tggtagcgaacaaatccaagtcggtgtgatggcgtagaagtttgccaaggtttaataata	3703
Qy	2955	gggttaattcaagcactggcattgatggcacagcgttatccaaaagtcaaatgtgctt	3014
Db	3704	tggcgttgtaggtgcctggcaatgatggcacaactgcatalccagagatgataatgtgctt	3763
Qy	3015	tactgggcctaaatgggtccaactgtataccccaaaaacccacacttaacaaagacaagcttaa	3074
Db	3764	tactggcactaaatgggtccaactgtataaagacaaacccacactaagcaaaagac-----	3815
Qy	3075	agtggttgaagttgaattaccaacactggcattaacgcaggtggttaaaaaagatatccaa	3134
Db	3816	-----ggcattaacgcaggtggttaaaagattaccaa	3847
Qy	3135	cattcaatcaggtgatattaccocaaacagcaatgatctgtgcagcgttcggttata	3194
Db	3848	cattcaatcaggttgagattgccocaaacagccatgatctgtgcagcggcgaagattata	3907
Qy	3195	tgattaaaaacccgaactgaaagcaaaatcaacagtgctgtctaaacacgacacaaactc	3254
Db	3908	tgattaaaaacccgaactgaaacaaaaatcagcagtagctgccaaaaacgcacaaaaactc	3967
Qy	3255	attacagtaattctcagtagcagatgaacaaagttaatcactttacggttagtaaccttta	3314
Db	3968	attacagtaattctcagtagcagatgaacaaagttaataactttacggttagtaaccttta	4027
Qy	3315	ctccagttatgacactc aaagacctctatgtctatccactttgcagtgaaaacgcat	3374
Db	4028	ctccagttatgacactc aaagacctctatgtctatccactttgcagtgaaaacgcat	4087
Qy	3375	taccacaaagttaaaataaagggttggtgcgttggtggcattgacccaaaccaaaggcttaac	3434
Db	4088	taccacaaagttaaaataaagggttggtgcgttggtggcattgacccaaaccaaaggcttaac	4147
Qy	3435	cacgcctaaagctgacogtgggttaataatggcaagdcattgtctatgcagataaaga	3494
Db	4148	cacgcctaaagctgacogtgggttaataatggcaagdcattgtctatgcagataaaga	4207
Qy	3495	tggctcaaaaaatccatcacaggactaagaacacttagctaatgttaccata	3548
Db	4208	tggctcaaaaaatccatcacaggactaagaacacttagctaatgttaccata	4267
Qy	3549	tggctcaggcacgcactaagccaaagggccttgccaatgacacccgacaaaacccgtgcgc	3608
Db	4268	tagcgttcgcacccacagacagggcgaatataatcaaaagcgaagcaaaaacccgtgcgc	4327
Qy	3609	cagcatttggtgatgtgctaaacgcaggccttcaacttgcgaagcaatggtgaagcgttga	3668
Db	4328	cagcatttggtgatgtgctaaacgcaggccttcaacttgcgaagcaatggtgaagcgttga	4387
Qy	3669	ctttgtctcaacttatgacactgttgactttatcgatggcaatgcacaccgcgttaaggt	3728
Db	4388	ctttgtctcaacttatgacaccgtcaactttcccgatggcaatgcacaccgcgttaaggt	4447
Qy	3729	gacctatgatgacacagcaaaaacccagtaagtgctctaatgtcataatgtcgataataa	3788
Db	4448	gacctatgatgacacagcaaaaacccagtaagtgctctaatgtcataatgtcgataataa	4507
Qy	3789	aacctattgaagtgcacaaagtataaaaaacttgccgttcaaaacccacacactgcacaaaac	3848

[illegible]

b	1136	aggtgaggtagtagtagtagtactacagtgaaattatggatgatagttaaoccttaacc	1195
y	1137	-----caacaacaaatgcgaagcagacagaacaaacccgtctatggcactgatggctaagt	1390
b	1196	agccaatacacaggcagtcbaagcagaacgaacccgtctatggcgtaatggggggaagt	1255
y	1391	ttaicgataaatcttaatacgcgcactttgaa---gatactactcgtatcatacccagaagataaaa	1447
b	1256	ttaataataatgcagaacaaacagcagcaaatcggcactactcgtattaccagagataaaa	1315
y	1448	ttggttttagcaataaagctggtagttgatgaaaaaacacttatcttgtataaagaca	1507
b	1316	ttggctttgctcgagatggtgatgttgatgaaaaacaagcaccatatattggataaaaaac	1375
y	1508	agctaaaaagtgg-----	1520
b	1376	aacttaagtgggtagttggttgcattaccatagacaatggcattgatgcaggtataaaaa	1435
y	1521	-----caacagcaccttaaacacagtggtgctgaactgttaataaacacca	1564
b	1436	agatcagtaacttgcgaagtagcagtgtcaacgtagcggttaccatcgacaagctca	1495
y	1565	ttggtgtagcaataaacaacaaatcccaagtcggtgctgatggcattaaatttgcgagatgtga	1624
b	1496	aagccgccaaagctactttaacgcgagcgctggcagtcagttcacacctactagaaatat	1555
y	1625	atgttaatgtatacaaatgcgcgcaaaatcggcactactcgtattaccgaagagagaaattg	1684
b	1556	cagttgctgaagatggtgcaatgttacgcgcccaacttaacaacttggcgtgaaaaacca	1615
y	1685	gctttgctgctgctgatggtaagttgataa-----	1715
b	1616	ccgagcttaacagtgatggcactagtagtaaaattagttgaaggtgtagtggtagcaaca	1675
y	1716	-----aaagtcaccatttggataaaaaacaacttc	1747
b	1676	atagcttagttaccgcgcaaatcttggcaagctatctaaatgaagtcaatcgacggctg	1735
y	1748	aagtgggtgggtgttaaaattaccaaaagacagtggtgcattaatcaggtgatcaaaagatca	1807
b	1736	acagtgctctaaaagctttaccggttaaagaagaacagatgatgacccaacgctatca	1795
y	1808	gtaatgttaagatgcacgcgagacacagatcagtcacttataaacagctttaacaag	1867
b	1796	ccgtggctcaagatcacacaacaaaatgccggcgagtcagcttcataaaactcaaagta	1855
y	1868	tccaacaagacgcgcagcgtgcccataca-----agctctctatctgtgatg	1915
b	1856	aaaacggtcaacggttgctaccacaaaagatgtagcgttacctttgggcttagccaag	1915
y	1916	aaaaaggtcaggaaattacgattagtaaacctgtattctaatggtataccccaaatacct	1975
b	1916	atagcggctcgaccatttggcaaaagcaccctaaacaacagatggcttgactgttaaagata	1975
y	1976	ttgagaccatcacctttgcaggtgaaaacggcatcagdtacgcaatgacatagccaaag	2035
b	1976	ccaacgaacaaatccaagtgcgtgtaatggcatataattactaatgtgaatggttagta	2035
y	2036	gtaaaqtaaaag---ttggtattgacccaatccaatggtctccacacgcttaagctgaccg	2092
b	2036	atccaggtactggcatttgcataacccgtcgcattaccagagataaaatttggctttgctg	2095
y	2093	tgsggtagcataaagatggtataaacatcaatttggttatttggaag-----	2137
b	2096	gttctgatggtgcagttgatatacaacaaacacttatcttgatacaagacaagctacaagttg	2155
y	2138	-----tggtacggtlaacgacacacacaaacatcatltagaggatgtt	2179
b	2156	gcaatgttaagattaccaacactggcattacgcgaggtggtggttaagccatcacagggctgt	2215
y	2180	ccccacactgcttagcattaccaattgcaggtggcgtagcaccacacagacaggggcaata	2239

Db	2216	cccaacactgctagcattgccgataccaagtagc---cgcaacatagaactgggcaata	2272
Qy	2240	caatcacacgagcgaagacaaatccaaagccgcagtagctggtatataataatacacg	2299
Db	2273	caat---ccaagacaagaacaaaccacgcgtgcgaactaaatgatataataatacacg	2329
Qy	2300	gctttaaccctaaaaataatagcaactccgttggcttggcttccacttatataacactggtg	2359
Db	2330	gctttaaccctaaaaataataacaacccattgacttggcttccacttatgacattggtg	2389
Qy	2360	actttatogtaggcaatgcccacacgcgctaagtaactcacgtgaaacccaatacaaca	2419
Db	2390	actttgccaatggcaatgcccacacgcgccaagtaaccatgataccgctaacaacaaca	2449
Qy	2420	gtaaagtaacttatgattgcaatggtagtagagaaaaccattgaactcacagcgcataatg	2479
Db	2450	gtaaagtggtatagatggaatggtagatatacaaacattctctcaaacgcaactggtg	2509
Qy	2480	gcaagacaaaaaataatggcgtaaaaaccacacactgacccacaacaatgctaaaggta	2539
Db	2510	acaa---taaaaaacttggctcaaaaccacacacactgcaaaaaacaagtctaatggta	2566
Qy	2540	aagcaac-----caactttagtcacccgataacgtagtgccttgttaacgccaaag	2590
Db	2567	atcagcaactaaactttaatgttaactctagtatgaagtgccttgttaacgccaaag	2626
Qy	2591	acatgcgcgaaaattaaacacccctagccaaggaattcacacacaaagggcacagcag	2650
Db	2627	acatgcgcgaaaattaaacacccctagccaaggaattcacacacaaagggcacagcag	2686
Qy	2651	acacgcctctacaaacctttaaagtcaaaaaagcggtagca-----actgagtacgcgaaa	2704
Db	2687	acacgcctctacaaacctttaaagtcaaaaaaggtagatgaaaaataataatgctgatgcg	2746
Qy	2705	ccatcacctgggttaaaagtgtacaaaaacggcaagacgcgtcaacactctaaaaactca	2764
Db	2747	ccaacgccatcacccgtgggtcaaaagacgcgaaataatcaagtcacacccctaaactca	2806
Qy	2765	aagtgaaaaacggtctaaacggttgctaccaaataagatggtagctgttaccttggcatta	2824
Db	2807	aagtgaaaaacggtcttaataataaaacgcgcaaaaaatggtagctgttaccttggcatta	2866
Qy	2825	acacccaaagcggctcttaaacgcgcgcacagcacacactctaaacaagaatggctgtcta	2884
Db	2867	acacccaaagcggctcttaaacgcgcgcgcaaaagcacc---ctaaacgcggtggctgtgcta	2923
Qy	2885	ttaaaaaccgcgtagtcaacgaacaaatccaaagtcggtgctgtagctggcgtgaagttgcga	2944
Db	2924	ttaaaaaccgcactggcagcgcaaaatccaagtgcggtgctgtagctggcgtgaagttgcga	2983
Qy	2945	aggt---tgataaagggtaattcaagcactggcaattgatggcacaagccgtatccacaaag	3001
Db	2984	aggttaataataatgggtgttgtaggtgctggcaattgatggcacaactgcgcatcacagag	3043
Qy	3002	atcaaatggctttactggggctaaatggctccacttgataccaccaaaacccaccctaaacca	3061
Db	3044	atgaattggctttactcgggactaaatggctccacttgataaaaagcaaaacccaccctaaagca	3103
Qy	3062	aagcaacgcttaaatgggtgaagttgaaattaccaacactggcattaacgcaggtggtta	3121
Db	3104	aagac-----ggcataacgcagatgggtta	3127
Qy	3122	aaaagattaccacaattcaatcaagttgattattacccaaacagcaatgatgtgtgacag	3181
Db	3128	aaaagattaccacaattcaatcaagttgattgctcccaaaaacgcaatgatgtgtgacag	3187
Qy	3182	gcggctcggtttatgatttataaaacccgaacttgaagcaaaaatacagatgctgtctaaaaa	3241
Db	3188	gcggcgaagatttatgatttataaaacccgaacttgaaaacaaaatacagactactgcccacaa	3247
Qy	3242	cagcaaaaactcattacagaattctcagtagcagatgaacaaaggtcaatcaactttcacgg	3301
Db	3248	cagcaaaaactcattacagaattctcagtagcagatgaacaaaggtcaatcaactttcacgg	3307

Qy	3302	ttagtaaccttactcagttatgacacctcaaaagacctctgatgtcatcaacctttgcag	3361
Db	3308	ttagtaaccttactcagttatgacacctcaaaagacctctgatgtcatcacotttcag	3367
Qy	3362	gtgaaaacggcattaccaccaaggtaataaagaagtgttgctgctgtggcgcatlgaccaaa	3421
Db	3368	gtgaaaacggcattaccaccaaggtaataaagaagtgttgctgctgtggcgcatlgaccaaa	3427
Qy	3422	ccaaggctttaaccacgcctaagctgacogtgggttaaataaataggccaaaggcatgtca	3481
Db	3428	ccaaggctlaaacacgcctaagctgacogtgggttaaataaataggccaaaggcatgtca	3487
Qy	3482	ttgaacgtaaaagtgttcaaaaaataccatacagaagctaagcaaacactctagctaaattga	3541
Db	3488	ttgacagccaaaatggttcaaaaataccatacagaagctaagcaaacactctagctaaattga	3547
Qy	3542	ccaatga-----tggctgacgacacgcactaagccaagggtcttgccaaatgacaccogca	3595
Db	3548	ccaatgataaagtagcgttacgccacacaagaacaggccaatatataatcaaagacgaagaca	3607
Qy	3596	aaacccgtgcgcacgacatttgtgtatgtgtctaaacycagggtttaocttgcgaaggcaatg	3655
Db	3608	aaacccgtgcgcacgacatttgtgtatgtgtctaaacycagggtttaocttgcgaaggcaatg	3667
Qy	3656	gtgaagcggittgactttgtctccacttatgacacttgttgaactttatcgaatggcaatgccca	3715
Db	3668	gtgaagcggittgactttgtctccacttatgacacttgttgaactttatcgaatggcaatgccca	3727
Qy	3716	ccacgcgtaaagtgacactatgatgacacaagcaaaacacagtaaaagtggctctatgatgtca	3775
Db	3728	ccacgcgtaaagtgacactatgatgacacaagcaaaacacagtaaaagtggctctatgatgtca	3787
Qy	3776	atgtggataataaaacocattgaagtgcgaagtgtataaaaaacttggcgtcaaaacaccaca	3835
Db	3788	atgtggatataaacocattgaagttaaa--gataaaaaacttggcgtaaaaaacaccaca	3844
Qy	3836	cactgacaaaacaagtgtcaatgtgtaatgca----accaaatttiagtgcgcgcgatggcg	3892
Db	3845	cattgacccgtactgtgcacaggctgtcaataaatttgccttaagcaatacgaactacttggcg	3904
Qy	3893	atgcctctgttaagccagtgatatgcgcacccatctataacttggcttggcgacatcc	3952
Db	3905	atgcctcttcaaggccagtgatatcgttgcctatctaaacaccttatctggcgacatcc	3964
Qy	3953	aaacgcgcaaaaggggcaagcccagaagcagctcagcaagctatgtggatgctgatggca	4012
Db	3965	aaactgcgcaaaaggggcaagcccagcgacaaactcagcaggctatgtggatgctgatggca	4024
Qy	4013	acaaggtcalctatgacagtaccgataagaagctactlcaagtcaatgacaagggtcaag	4072
Db	4025	ataaggltca tctatgacaglaccgataacaagctactlcaagccaaaatgtatggccag	4084
Qy	4073	tggcaaaaaaagaagttgccccagaacaaactggttcgcccgaagcccaaacccccagatg	4132
Db	4085	ttgataaaacaaaagaagtgtccccagaacaaactggttcgcccgaagcccaaacccccagatg	4144
Qy	4133	gcacatttgttc aaatgtaatgttcaaatcagttcaatacagaagcagaagtaaatgatgccca	4192
Db	4145	gcacatttgttc aaatgtaatgttcaaatcagttcaatacagaagcagaagtaaatgatgccca	4204
Qy	4193	ataaaagcaggcatcaatgaaagcaaacgcctttatcaaaagggcttgaaaagccgcgcca	4252
Db	4205	ataaaagcaggcatcaatgaaagcaaacgccttttgaagggtctgaaaagccgcctt	4264
Qy	4253	aagacacaaaaacaaaaacgcgcagtaacttgttgggtgatttaaatgcccgtttgcccaaa	4312
Db	4265	ctgatacaaaaacaaaaacgcgcagtaacttgttgggtgatttaaatgcccgtttgcccaaa	4324
Qy	4313	cacgcgtgacctttgacggggatacagggcacaacggctataaaaacttgggcgagactttga	4372
Db	4325	cacgcgtgacctttgacggggatacagggcacaacggctataaaaacttgggcgagactttga	4384

Qy	4373	ccatcaaaaggtggcgaacacagacacacaaataaagctaacgcgataataacatcggttggtgtag	4433
Db	4385	ccatcaaaaggtggcgaacacagacacacaaataaagctaacgcgataataacatcggttggtgtag	4444
Qy	4433	cagtgactgatggcttcactgtcaaaacttgcacaaagacctaaccaatcttaacagcgtta	4492
Db	4445	cagtgactgatggcttcactgtcaaaacttgcacaaagacctaaccaatcttaacagcgtta	4504
Qy	4493	atgcaggtggcacacagaattgatgaaaaaggcatctctttttagacgacaaacggtcaag	4552
Db	4505	atgcaggtggcacacagaattgatgaaaaaggcgctctttttagactcaagcggctcaag	4564
Qy	4553	ccaaagcaaacacacccctgtctaaagtgcacaaatggcctggacctgggtggcaaacgcgtca	4612
Db	4565	ccaaagcaaacacacccctgtctaaagtgcacaaatggcctggacctgggtgggc	4614
Qy	4613	gtaacatcgggtgcagctgtgtatgataacgatgcggtgaaccttaagcagtttaataag	4672
Db	4615	-----	4614
Qy	4673	ttgcacaaaacggtcaacaaactaaacacaaacaaagtaactcaggtgcgtcattaccccttg	4732
Db	4615	-----	4614
Qy	4733	tggtaaccgatgccaatggcaagcccatcaatggcacogatggcaagccccaaaagcca	4792
Db	4615	-----	4614
Qy	4793	tcaaggcgccgatggttaaatactatcacgcccacacgcgcaacggcgtaacctgtggacaaa	4852
Db	4615	-----	4614
Qy	4853	atggcaagcccatcacogatgcggacaaacttgcacaaatctggcagctcatgtgcaaacccc	4912
Db	4615	-----	4614
Qy	4913	ttgatgcaggtcatcaagtgggtggcaagcctaggcggcaactcagatgcacatcacccctaa	4972
Db	4615	-----	4614
Qy	4973	ccaacatcaagtcacacttggccacaaattgacacacacaaacacaggttaatgccaatgcag	5032
Db	4615	-----	4614
Qy	5033	ggcaagcccaaaagtctgccagcctatcacgacgacgacagcaaaagtaatgctgccagtgca	5092
Db	4615	-----	4614
Qy	5093	aagatgtctaaatgtaggctttaacttgcagaccaatcacaaatcaagtggactttgtca	5152
Db	4615	-----	4614
Qy	5153	aagcctatgatacgttcaactttgtcaatgggtacaggtgcggacatacacaagcgtgcgta	5212
Db	4615	-----	4614
Qy	5213	gtgctgatggcaogatgtagtaacatcacccgtcaacacccgcttagcagcgcgatgatg	5272
Db	4615	-----	4614
Qy	5273	atggcaatgtgcttatcaaaagccaaagatggtaagtcttacaaagcacagcacctcagc	5332
Db	4615	-----	4614
Qy	5333	caacgcgctcactaaaagcaggcaaatcagccagtgatgcacaaactccaactggtctaa	5392
Db	4615	-----	4614
Qy	5393	gcctgttaaccocaaatgctggttaaaggcagtcagtcagtcggtggtctctaataact	5452
Db	4615	-----	4614
Qy	5453	tatcaaaagcggattattaaatccaaagatgggtacaaactactaccacagtaagctctgatg	5512

[illegible]

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2001, 12:57:17 ; Search time 454.28 Seconds
(without alignments)
2892.925 Million cell updates/sec

Title: US-09-361-619-10
Perfect score: 6942
Sequence: 1 atgaatcacatctataaagt.....ttggtcagggttttcacttt 6942

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2580.6	37.2	6973	1	US-08-478-370-1
2	2382.4	34.3	9542	4	US-08-968-685A-9
3	83.6	1.2	2037	4	US-08-913-942-14
4	82.8	1.2	3294	1	US-08-409-995-1
5	82.8	1.2	3294	3	US-08-685-467-1
6	82.8	1.2	3294	4	US-08-913-942-1
7	76.4	1.1	7291	4	US-08-913-942-3
8	69.8	1.0	5738	1	US-08-409-995-3
9	69.8	1.0	5738	3	US-08-685-467-3
10	58.4	0.8	60	1	US-08-478-370-4
11	56.8	0.8	1797	4	US-09-377-155-12
12	56.8	0.8	1797	4	US-09-377-155-4
13	55.8	0.8	7218	1	US-08-232-463-14
14	55.2	0.8	1770	4	US-09-377-155-18
15	55.2	0.8	1776	4	US-09-377-155-10
16	55.2	0.8	1776	4	US-09-377-155-20
17	55.2	0.8	1779	4	US-09-377-155-3
18	55.2	0.8	2308	4	US-09-377-155-1
19	53.6	0.8	1785	4	US-09-377-155-6
20	53.6	0.8	1785	4	US-09-377-155-8
21	53.6	0.8	1800	4	US-09-377-155-14
22	52	0.7	1779	4	US-09-377-155-16
23	48.2	0.7	15222	2	US-08-801-898A-23
24	48.2	0.7	15222	4	US-08-962-690-12
25	48.2	0.7	15223	2	US-08-892-403A-1
26	43.4	0.6	894	2	US-08-467-963C-28
27	43.4	0.6	894	2	US-08-838-189D-28

28	43.4	0.6	894	3	US-08-852-344D-28	Sequence 28, Appl
29	43.4	0.6	894	3	US-08-344-639E-28	Sequence 28, Appl
30	43.4	0.6	920	2	US-08-467-963C-7	Sequence 7, Appl
31	43.4	0.6	920	2	US-08-838-189D-7	Sequence 7, Appl
32	43.4	0.6	920	3	US-08-852-344D-7	Sequence 7, Appl
33	43.4	0.6	920	3	US-08-344-639E-7	Sequence 7, Appl
34	43.4	0.6	920	4	US-08-467-969A-7	Sequence 7, Appl
35	43.4	0.6	920	4	US-08-467-961A-7	Sequence 7, Appl
36	43.4	0.6	920	4	US-08-001-554A-7	Sequence 7, Appl
37	42.2	0.6	1236	2	US-08-741-134-5	Sequence 5, Appl
38	42	0.6	3489	2	US-08-728-323A-1	Sequence 1, Appl
c 39	42	0.6	32207	2	US-08-770-379-20	Sequence 20, Appl
c 40	42	0.6	32207	4	US-08-757-669A-20	Sequence 20, Appl
41	41.2	0.6	390	4	US-09-197-649-7	Sequence 7, Appl
42	38.4	0.6	1056	4	US-09-363-189B-3	Sequence 3, Appl
43	38	0.5	1760	1	US-08-413-118-117	Sequence 117, App
44	38	0.5	1760	3	US-08-473-446-117	Sequence 117, App
45	38	0.5	2280	1	US-08-220-151-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-478-370-1
; Sequence 1, Application US/08478370
; Patent No. 5808024
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/478,370
; APPLICATION NUMBER: US/08/478,370
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-502
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-370-1

Query Match 37.2%; Score 2580.6; DB 1; Length 6973;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 4803; Conservative 0; Mismatches 1104; Indels 1270; Gaps 29;
Qy 1 atgaatcacatctataaagtctcttttaacaaagccacaggttatgtgcgtggca 60
|||||

Db 542 ATGAATCACAATCTAATAAGTCATCTTTAAACAAAGCCACAGGCACATTTATGGCAGTGCCA 601
QY 61 gaggcgcaaatccacagc-9gagggagtagcagtagtaccacagacaggtggcgag 119
Db 602 GAGTAGCCCAATCCACAGCAGCGGGGGGGGTAGCTGTCTACAGGCGAAGTGGCAG 661
QY 120 ctctctgtcaccgcctgactcggtgtgcccacgctcgtctatccttagtgcgtgagac 179
Db 662 TGTATGCACCTGTGAGCTTTGCCCGTATTGCCGCGCTGCTGTCTCGTATCGGTGCAAC 721
QY 180 gctcaatggcagtgcttattgctcaaa-----ataatagcaagatcgcatattgt- 228
Db 722 GCTCAGTGGCAGTGTATGCTCAAAAAAAGATACCAAAACATATCGCAATTTGGTGAACA 781
QY 229 ----accacaggaacaatgcaatgctcggctagcaatgaagcatcattgctattgg 284
Db 782 AAACAGCCAAAGCGCTCAGGCACTGCCAAGGGGAGCGGTGATCGAGCCATTTGCTATTGG 841
QY 285 tagtctgtcaagggcacatgcccgaatcaagctattgcta cgggtggtagcaaacacagatcc 344
Db 842 TGAANAATGCTTAACGCACAGCGGGGTCAAGCCATCGGCATCGGTAGTAGTA 898
QY 345 tgcgtaatacagcgggcgaatcaagaagcaggttccacgcgcaaaaggttaaagagtcctatgc 404
Db 899 TGTCAATCGAAGCAGTTTGGATAAGATAGGTACCGATGCTACGGGTCAAGATGCCATCGC 958
QY 405 catcgggtggtgtagtactggtgaggtgtagtgcctcgtatgcatcattggttagtgcatt 464
Db 959 CATCGGTGGTGTATAAGGCTAGTGTGTGCTGCTCGATTGCCATCGGTAGTAGTACAT 1018
QY 465 atatttggatagggaatagcactaacttaataatacaaatgg---tcttcttagcactct 521
Db 1019 ACATTTGCTTGATCAGCATGATGATCTCTAAACATFCCGAAGGTACTCTGTATTACGATCT 1078
QY 522 tattcaaaaccatcacagattacgcccataacacagagctcaaatgggtctctc---gaata 578
Db 1079 TATTAAACCGCCATGCAGTATTAAAAAGAAATACGAAGTCTCAAAAGGATAATGATGATAA 1138
QY 579 tagacgcacagcagcagaagaacagccagctactgactgagtgagcagcagctatgcata 638
Db 1139 TAGAGGCACAAACCGCAACGCGACAGCGCAGTACTGCAAGTGGGAGCCATGTCATATGCACA 1198
QY 639 gggctatttgcacgccttgggtacacggtcacaacagctgaagggcaactattccttggc 698
Db 1199 GGGTCATTTTCCACGCTTTGGTACACGGGCAACAGCTAAAAGTGCCTATTCTTGGC 1258
QY 699 agtaggtcttacgcgcgaagccgaaggaatatacaatcgctattggttctaatgcaca 758
Db 1259 AGTGGGTCTTGGCGCCACAGCGCGAGGGCCAAATCTACAATCGCTATTGGTCTGATGCAAC 1318
QY 759 agctatcaattatggagcactagcccttgggtgcagatactcaggttgatttggattacgg 818
Db 1319 ATCTAGCTCGTTGGGAGCGATAGCCCTTGGTGCAGGTACTCGTCTCAGCTACAGGGCAG 1378
QY 819 tattgcccattggttagtgcactatccttaataataataataata---ataataa 875
Db 1379 TATTGGCCTAGGTCAGAGTCTGTGTCTACTCAGAGTGATAATAATTCTAGACGGGCTA 1438
QY 876 agcctatgtaccagaaggtlaattgggtccaaacataaaatcgcttaagccacgcgcaatgg 935
Db 1439 TACACCAAAATACCCAGGCATAGACCCCAAGTTTCAAGCCACCAATAATAACGAAGCGGG 1498
QY 936 ttattttccatttggtagtagcactatcgaagcgaataaataatcaatcggtgcaggtta 995
Db 1499 TCCACTTTTCCATTGGTAGTAACTCTATCAACGTAATAATCAATATGTGCGGTGAGGTGT 1558
QY 996 tgaagatcacgagtcggtcaatgtggcacagctaaagcgggtggagaatcctggcta---a 1052
Db 1559 TAATAAACCAGTCGGTCAATGTGGCACAGCTAGAAGCGGTGGTGAAGTGGGCTAAGGA 1618
QY 1053 gcgtaaatctacttttaagggtagatgaacggttactggtggttaagaaaaaactggcgca 1112
Db 1619 GCGTAGAATTAATTTTTCAGGGGTGATGATAACAGTACTGACGTAAAAAATAGGTTGGATAA 1678

QY 1113 gactttaaccattaaagtggtgagacccaagcgacagctaaaccgataataataaacat 1172
Db 1679 TACTTTAACTATTAAAGGTGGTG---CAGAGACCAACGCATTAAACCGATAATAATATCGG 1735
QY 1173 tgggtggttaacagataataataactgggtctgaaagctaaacttgtaaaaaaccctaaagcg 1232
Db 1736 TGTGGTAAAAAGAGGCTGATAATAGTGGTCTGAAAGTTAAACTTGTAAACCTTTAAACAA 1795
QY 1233 tcttgaacaagtttagcaccacaaacacctaaccgcccagcgagaaagttacggtag---gtag 1289
Db 1796 TCTTACTAGAGTGAATACAACTACATTAATGCCACAACACACAGTTAAGGTAGGTAGTAG 1855
QY 1290 tggtaataacacacgcgtgagctacaaagcggtggtttbaacctttacc-----caacaac 1343
Db 1856 TAGTAGTACTACAGCTGAATTTATTGAGTGTATGTTTAACTTTTACCACGCCCAATACAGG 1915
QY 1344 aaatgcaagcaagacaaacacccgtctatggcagctggtggttaagtttactgataaattc 1403
Db 1916 CAGTCAAGCACAGCAAGCAAAACCCGTCTATGGCGTTTAAATGGGGTGAAGTTTACTAATAATGC 1975
QY 1404 taatcagggcacttga---gatactactgatacacaagaataaaattggttttagcaa 1460
Db 1976 AGAAACACACAGCAGCAATCGGCACCTACTCGTATTACCAGAGATAAAATTTGGCTTCG 2035
QY 1461 taaagctgggtacagttgatgaaacaaacacccattcttctgataaagacaaagctaaag--- 1516
Db 2036 AGATGGTGTATTGATGAAAAAACAGCACCATATTGATATAAAACCACTTAAAGTGGG 2095
QY 1517 -----t 1517
Db 2096 TAGTGTTCCAATTTACCATAGACAAATGCCATTGTATGCAGGTAAATAAAAGATCATCATCT 2155
QY 1518 tggcaacagcaccctaaacaacggtggtctgacgtgtaataaacaacccattggtgtagcaa 1577
Db 2156 TGCCAAAGGTAGCAGTGTCAACGATGCGGTTACCATCGACAGCTCAAGGCCCAAGCC 2215
QY 1578 taaacaataacaaacggtgctgctgattgacattaaatttgcgagtgatggaattgtaattgctc 1637
Db 2216 TACTTTAAACGCGAGCGCTGGCATCAGTGCACACCTACTGAAATATCAGTTGATGTATA 2275
QY 1638 aaatgccgcgaatctggcactactcgtattaccgaaaggaagaaatttggcttctgctgac 1697
Db 2276 GAGTGGCAATTTTACCGGCCCAACTTACAACATTTGGCGTGAAACCCACCGAGCTTAACAG 2335
QY 1698 tgaatgtaaaagtgtgataa----- 1715
Db 2336 TGATGGCACTAGTGTATAAATTTAGTGTAAAGGTAGTGTACGAACAATAGCTTAGTTTAC 2395
QY 1716 -----aaagtcacccatatttggataaaaaaacacttcaagtggtggtgt 1760
Db 2396 CGCGCAACATTTGGCAAGCTATCTAAATGAAGTCAATCGACGGCTGACGTGCTCTACA 2455
QY 1761 taaattaccaaagacagtggtcatttaagtgcgtgatacaaaagatcagtaattgttaaaga 1820
Db 2456 AAGCTTTTACCCTTTAAAGAGAAGACAGCATGATGACGCCAACAGCTATCACCGTGGCTTAAAGA 2515
QY 1821 tgcacaggaacgacatcccgatgcgtcacttataaacagcgtttaaacaagtcaccaacaagcgc 1880
Db 2516 TACGACAAAAAATGCCGGCGCAGTCAGCATCTTTAAACCTCAAAAGGTAAAACCGGTCTAAC 2575
QY 1881 cgaagctgacctcaaa-----agcttctctatttcgtgtgtaaaaaaggttcagga 1928
Db 2576 GGTGTGTACCAAAAAGATAGTGTACGGTTACCTTTGGGCTTAGCCAAGATAGCGGTCTGAC 2635
QY 1929 atttaccattagtaacttcttatttatttatttatttatttatttatttatttatttatttattt 1988
Db 2636 CATTTGGCAAAAGCACCTTAAACACAGTGGCTTGACTGTTTAAAGATACCAACCAACAAAT 2695
QY 1989 ctttgcaggtgaaacacggccatcagtatcagcaatgacatagcacaaggttaaagtcctaaag- 2047
Db 2696 CCAAGTCGGTGTCTAATGGCATTTAAATTTTACTAATGTCAATGGTAGTAATCCAGGTACTGG 2755

D	b	4865	AATGAATGTCAAAATCAGTTCATTACAAAGAACAAAGTAATATGATGCCAATAAATAAGCAGG	4924
Q	y	4206	catcaatgaagacaacgctttatcaagggttgaaacgcgcgcaaaagacacacaaac	4265
D	b	4925	CATCAATGAAGACACGCGCTTGTTAAGGACTTGAAAAGCCGCTTCTGATAACAAAAC	4984
Q	y	4266	caaaaacgcgcgcagtaactgtgggtgatttaaatgcccgttcccaaaaacacgcgtgaactt	4325
D	b	4985	CAAAAACGCGCAGTAACGTGTGGGTGATTTAAATGCCGTTGCCCAACACCCGCTGACCTT	5044
Q	y	4326	tgcagggtacagcacaacgcgttaaaactggcgagactttgaccatcaaaaggtgg	4385
D	b	5045	TGCAGGGGTACAGGCACACACGCGCTAANAACCTGGCGGAGACTTTGACCATCAAAAGGTGG	5104
Q	y	4386	gcaacagacacccaataagctaacccgataataacatcctggtgtgtagcaggtactgtatgg	4445
D	b	5105	GCAAAACAGACCAATAAGCTAACCGGATAATAACATCGTGTGTGGTAGCAGGTACTGTATGG	5164
Q	y	4446	cttcactgtcaaaacttgcaaaagacctaaccaatcttaacagcgttaatgcaaggtggcac	4505
D	b	5165	CTTCACTGTCAAACTTGCCCAAGACCTAACCAATCTTAACAGCGTTAATGCAGGTGGCAC	5224
Q	y	4506	cgaattgatgaaaaaggcatctctttgtagacgcacaacgcgtcgaagccaaagcaaacac	4565
D	b	5225	CAAAATTGATGACAAAGGCGTGTCTTTGTAGACTCAAGCGGTCAAGGCCAAAGCAAAACAC	5284
Q	y	4566	ccctgtcgttaagtgcacaatgggctggacctggtggcaaacgcacatcagtaacatcgttgc	4625
D	b	5285	CCCTGTGCTAAGTGCCAAATGGGCTGGACCTGGGTGGC-----	5321
Q	y	4626	agctgtgatgataacgaatgcggtgaactttaagcagtttaagtaagtgtgcacaaacaggt	4685
D	b	5322	-----	5321
Q	y	4686	caacaacctaaacaacccaaagtaactcagtgcggtcattaccctttgtgttaaccgatgc	4745
D	b	5322	-----	5321
Q	y	4746	caatggcaagcccatcaatggcaccgattggcaagcccccaaaagccatcaaggggcgccga	4805
D	b	5322	-----	5321
Q	y	4806	tggtaatactatacgcgccaacccaacgcgcgtacctgtggcaaaagattggcaagcccat	4865
D	b	5322	-----	5321
Q	y	4866	caccgatgcggacaaaacttgccaatctggcagctcatggcaaccccttgatgcaggtcca	4925
D	b	5322	-----	5321
Q	y	4926	tcaagtgtggcaagcctagggcaactcagatgcctatcccttaacccaacatcaagtc	4985
D	b	5322	-----	5321
Q	y	4986	cactttgccacaaattgacacacaaaacacaggttaattgccaatgcaagggcagcccaag	5045
D	b	5322	-----	5321
Q	y	5046	tctgccagcctaatacagcagcagcaaaagtaattgctgcagtgtaaaagatgtgetaaa	5105
D	b	5322	-----	5321
Q	y	5106	tgtaggctttaacttgagaccacaatacaatcaaatggaactttgtcaaaagcctatgatac	5165
D	b	5322	-----	5321
Q	y	5166	cgtcaacttgtlcaatggtacaggtgcgcagatcacaaagcgtgcgtagtgatggcac	5225
D	b	5322	-----	5321
Q	y	5226	gatgagtaacatcacccgtcaacacgcgccttagcagccagatgatgatggaatgtgct	5285
D	b	5322	-----	5321

Q	y	5286	tatcaaaagccaaagatggtaagttctacaagaagagacgacctcatgccaaaacggtcact	5345
D	b	5322	-----	5321
Q	y	5346	aaaagcaggcaaatcagccagtgatgccaaaactccaaactggctaaagccttgttaacccc	5405
D	b	5322	-----	5321
Q	y	5406	caatgctggtaaaggcagtagacagcgatgcagtggtccttaataaacttatcaaaaagcggt	5465
D	b	5322	-----	5321
Q	y	5466	atttaaatccaaagatggtatacactactaccagtaagctctgatggcatcagtatccca	5525
D	b	5322	-----	5321
Q	y	5526	aggcaaaagataaacagcgatcacccctaaagcaaaagatgggctgaatgtaggcggtgaaggt	5585
D	b	5322	-----	5326
Q	y	5586	catcagcaatgtgggtaaaggcacaagaagacacgcgcgtgccaatgtacaacagtttaaa	5645
D	b	5327	CATCAGTAAATGTGGCAAGGCACAAAGATACCGACGCTGCCAATGTACAACAGTTAAA	5386
Q	y	5646	cgaagtacgcgaactgtttgggtccttggttaatgctggtaataatgataaacgctgcgcgaatca	5705
D	b	5387	GAAAGTACGCCAACTTGTGGGTCTTGGTAAATGCTGGTAAATGATAACGCTGACGGCAATCA	5446
Q	y	5706	ggtaaacattcccaactcaaaaaagacccaaatctcaggttcatctataaccgcactgt	5765
D	b	5447	GGTAAACATTCGCCGACATCAAAAAGACCCAAATTCAGGTTTCATCTAATCCGCACGTGT	5506
Q	y	5766	catcaaaagcagcgactgttggtggtggttaaggttaataaacgataccccgaaaaacttgcac	5825
D	b	5507	CATCAAAACAGCACGCGTACTTGGCGGTAAAGGTAAATTAACGATACCCGAAAACTTGCAC	5566
Q	y	5826	tgggtgtacaagtggcggtggataaaagacgcgaacgtaacgctaacgcgctttaagcgaatgt	5885
D	b	5567	TGGTGGTATACAAGTGGCGGTGGATAAAGACGGCAACCGCTAACGGCGATTTTAAGCAATGT	5626
Q	y	5886	ttgggtcaaaaaccccaaaagatggcagcaaaaaagccctgtctgcaccttataacgcgcgc	5945
D	b	5627	TTGGGTCAAAACCCCAAAAGATGGCAGCAAAAAGCCCTGTCTGCCACTTATTAACGCCGC	5686
Q	y	5946	aggtcagacccaacttttgaccaacaaacccgcagaaacatttgcagataaataagcaaca	6005
D	b	5687	AGGTGAGACCAACTATTGTACCAACAAACCCCGCAAGCCATTGACAGAAATTAATGAACA	5746
Q	y	6006	aggtatccgcttcttccatgtcaacgattggcaatcaaaagcctgtggtaacaggcgctaa	6065
D	b	5747	AGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCGTGGTACAAAGGCGCTAA	5806
Q	y	6066	cggcattgactcaagtgcctcaggaagcactcagtggcgataagtttccagggccaaggc	6125
D	b	5807	CGGCATTGACTCAAAGTGCCTCAGGCAAGCACCTCAGTGGCGATAGGTTCACAGGCCAAGGC	5866
Q	y	6126	agatggtgaagccgcgcttgccataggcagacaaaccccaagcagcaacccaactccatgcgc	6185
D	b	5867	AGATGTGAAGCCGCCGCTTGCCATAGGCAGACAAACCCCAAGCAGGCAACCAATCATCGC	5926
Q	y	6186	catcgtgataacgcacaagccacggcgatcaatccatcgcacatcgggtacagggcaaatgt	6245
D	b	5927	CATCGGTGATACGCACAAGCCACGGCGCATCAATTCATTCGCCATCGGTACAGGCAATGT	5986
Q	y	6246	ggtaaacaggttaagcacacttgggtgcaccccaagcactgttaagcctgataaacag	6305
D	b	5987	GGTAGCAGGTAAACACTCTGGTGCCATCGCGACCCCAAGCACTGTTAAGGCTGATTAACAG	6046
Q	y	6306	ttacagtggtggttaataacaacccagtttatcagatgcacatcagacagcagtgctttggtgt	6365
D	b	6047	TTACAGTGTGGGTAAATAACAACACCGTTTACCGATGCCATCAAAACCGATGCTCTTTGGTGT	6106

Thu Sep 13 14:17:56 2001

1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2819 2820 2821 2822 2823 2824 2825 2826 2827 2828 2829 2830 2831 2832 2833 2834 2835 2836 2837 2838 2839 2840 2841 2842 2843 2844 2845 2846 2847 2848 2849 2850 2851 2852 2853 2854 2855 2856 2857 2858 2859 2860 2861 2862 2863 2864 2865 2866 2867 2868 2869 2870 2871 2872 2873 2874 2875 2876 2877 2878 2879 2880 2881 2882 2883 2884 2885 2886 2887 2888 2889 2890 2891 2892 2893 2894 2895 2896 2897 2898 2899 2900 2901 2902 2903 2904 2905 2906 2907 2908 2909 2910 2911 2912 2913 2914 2915 2916 2917 2918 2919 2920 2921 2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2934 2935 2936 2937 2938 2939 2940 2941 2942 2943 2944 2945 2946 2947 2948 2949 2950 2951 2952 2953 2954 2955 2956 2957 2958 2959 2960 2961 2962 2963 2964 2965 2966 2967 2968 2969 2970 2971 2972 2973 2974 2975 2976 2977 2978 2979 2980 2981 2982 2983 2984 2985 2986 2987 2988 2989 2990 2991 2992 2993 2994 2995 2996 2997 2998 2999 3000 3001 3002 3003 3004 3005 3006 3007 3008 3009 3010 3011 3012 3013 3014 3015 3016 3017 3018 3019 3020 3021 3022 3023 3024 3025 3026 3027 3028 3029 3030 3031 3032 3033 3034 3035 3036 3037 3038 3039 3040 3041 3042 3043 3044 3045 3046 3047 3048 3049 3050 3051 3052 3053 3054 3055 3056 3057 3058 3059 3060 3061 3062 3063 3064 3065 3066 3067 3068 3069 3070 3071 3072 3073 3074 3075 3076 3077 3078 3079 3080 3081 3082 3083 3084 3085 3086 3087 3088 3089 3090 3091 3092 3093 3094 3095 3096 3097 3098 3099 3100 3101 3102 3103 3104 3105 3106 3107 3108 3109 3110 3111 3112 3113 3114 3115 3116 3117 3118 3119 3120 3121 3122 3123 3124 3125 3126 3127 3128 3129 3130 3131 3132 3133 3134 3135 3136 3137 3138 3139 3140 3141 3142 3143 3144 3145 3146 3147 3148 3149 3150 3151 3152 3153 3154 3155 3156 3157 3158 3159 3160 3161 3162 3163 3164 3165 3166 3167 3168 3169 3170 3171 3172 3173 3174 3175 3176 3177 3178 3179 3180 3181 3182 3183 3184 3185 3186 3187 3188 3189 3190 3191 3192 3193 3194 3195 3196 3197 3198 3199 3200 3201 3202 3203 3204 3205 3206 3207 3208 3209 3210 3211 3212 3213 3214 3215 3216 3217 3218 3219 3220 3221 3222 3223 3224 3225 3226 3227 3228 3229 3230 3231 3232 3233 3234 3235 3236 3237 3238 3239 3240 3241 3242 3243 3244 3245 3246 3247 3248 3249 3250 3251 3252 3253 3254 3255 3256 3257 3258 3259 3260 3261 3262 3263 3264 3265 3266 3267 3268 3269 3270 3271 3272 3273 3274 3275 3276 3277 3278 3279 3280 3281 3282 3283 3284 3285 3286 3287 3288 3289 3290 3291 3292 3293 3294 3295 3296 3297 3298 3299 3300 3301 3302 3303 3304 3305 3306 3307 3308 3309 3310 3311 3312 3313 3314 3315 3316 3317 3318 3319 3320 3321 3322 3323 3324 3325 3326 3327 3328 3329 3330 3331 3332 3333 3334 3335 3336 3337 3338 3339 3340 3341 3342 3343 3344 3345 3346 3347 3348 3349 3350 3351 3352 3353 3354 3355 3356 3357 3358 3359 3360 3361 3362 3363 3364 3365 3366 3367 3368 3369 3370 3371 3372 3373 3374 3375 3376 3377 3378 3379 3380 3381 3382 3383 3384 3385 3386 3387 3388 3389 3390 3391 3392 3393 3394 3395 3396 3397 3398 3399 3400 3401 3402 3403 3404 3405 3406 3407 3408 3409 3410 3411 3412 3413 3414 3415 3416 3417 3418 3419 3420 3421 3422 3423 3424 3425 3426 3427 3428 3429 3430 3431 3432 3433 3434 3435 3436 3437 3438 3439 3440 3441 3442 3443 3444 3445 3446 3447 3448 3449 3450 3451 3452 3453 3454 3455 3456 3457 3458 3459 3460 3461 3462 3463 3464 3465 3466 3467 3468 3469 3470 3471 3472 3473 3474 3475 3476 3477 3478 3479 3480 3481 3482 3483 3484 3485 3486 3487 3488 3489 3490 3491 3492 3493 3494 3495 3496 3497 3498 3499 3500 3501 3502 3503 3504 3505 3506 3507 3508 3509 3510 3511 3512 3513 3514 3515 3516 3517 3518 3519 3520 3521 3522 3523 3524 3525 3526 3527 3528 3529 3530 3531 3532 3533 3534 3535 3536 3537 3538 3539 3540 3541 3542 3543 3544 3545 3546 3547 3548 3549

Qy	2599	gaaatctaaacacccctagccaaaggaattcacaccaccaaaggaacagagacacgcgc	2658
Db	3242	GACAATCTAAACACCCCTAGCTGGTGAATTCGCAACACCAAGGCGCAGCAGACACCGCC	3301
Qy	2659	ctacaacaccttaagtcaaaaaa	2703
Db	3302	CTACAAACCTTTCAAGTCANNAAGTCAANAAGAAATGGTGATGATGATATGACGCTGAC	3361
Qy	2704	accataccctgggttaagatggtcacaaaaacggaagaccgtcaacactctaaaaactc	2763
Db	3362	ACCATCACCGCTGGGTAAGATG	3415
Qy	2764	aaagtgaaaaacgggtctaacgggttgctaccataaagaatggttacggttacctttggcatt	2823
Db	3416	AAAGGTATAAACCGGCTTGATATTCAAAACCAATAAAGATGTCAGGTTACTTTGGCATTT	3475
Qy	2824	aacacccaagcggtctaaagccggcgacagcacccactctaaacaaagatggctgtctct	2883
Db	3476	AACACCCAAAGCGGCTCTTAAGCGGCAACACACCACTCTTAACACACNAATGGCTGTCT	3535
Qy	2884	attaaaccccgctagtaacgaacaaatccaagtcggtctgctgtagtgcgtgaagttggcc	2943
Db	3536	ATTAANAACACCGCTGGTAACGAACAAATCCAAAGTCGGTGTGATGGCGTGAAGTTTGCC	3595
Qy	2944	aaggttgataaagggaattcaagcactggcattgtagcacaagccgatacaccaagaat	3003
Db	3596	AAGGTTAATAATGGTGTGTAGTGCTGGCATTTGATGCACAACCTGCCATTTACACAGAT	3655
Qy	3004	caaatggctttactgggctaaatggctcacttgatacccaaacccccaccactaaccaaa	3063
Db	3656	GAATTTGGCTTTGCTGGGACTAATGGCTCACTTGATATAAGACAAACCCACCTAAGCAAA	3715
Qy	3064	gacaagctaaagtgggtgaagttgaattacccaacactggcattaacagcaggttggtaaa	3123
Db	3716	GAC	3739
Qy	3124	aagattaccaaacattcaatcaggtgatatatacccaaaacagcaaatgatgctgtgacaggc	3183
Db	3740	AAGATTACCAACATTCAAATCAGTGAGATTGCCAAAACAGCAATGATGCTGTGACAGGC	3799
Qy	3184	ggtcggttttatgatctaaaaacccaacttgaagacaaaaatcaacagtgctgctctaaaaca	3243
Db	3800	GGCAAGATTTATGATTTAAACCGGAACCTTGAANACAAATCAGCAGTACTGTCCAAAACA	3859
Qy	3244	gcacaaactcatatacacgaattctcagtagcagatgaacaaggtaatcactttacggtt	3303
Db	3860	GCACAAAACTCATTACAGCAATTTCTCAGTAGCAGATGAACAAGGTAATAACTTTACGGTT	3919
Qy	3304	agtaacccttacctcagttatagacacctcaagacctctgatgcatcaaccttgcaggt	3363
Db	3920	AGTAACCCCTTACTCCAGTTATGACCTCAAGACCTCTGATGTATCAACCTTTTGCAAGT	3979
Qy	3364	gaaacggcattaccaccaaggttaataaagtgctgggtgctgtgggcattgaccaaac	3423
Db	3980	GAAACGGCATTAACCAACGAGTTAATAAAGTGCTGGTGGCTGGGCAATTGACCAACC	4039
Qy	3424	aaagccttaaccacgcctaaagctgacggtggtgtaataataatggaagggcattgtcatt	3483
Db	4040	AAAGGCTTAACACAGCCTTAAGCTGACCGTGGGTAATAATAATGCAAAAGGCAATGTGCATT	4099
Qy	3484	gacagtaagaatggtcaaaatccattcacagacactaaacacactctagtaattgttacc	3543
Db	4100	GACAGCCAAATGGTCAAAATACCATCAGGACCTAAGCACACTCTAGCTTAATGTTTACC	4159
Qy	3544	aatga-----tggtgcaggacacgcacttaagcccaagggcttgccaatgacacgcacaaa	3597
Db	4160	AATGATAAAGGTAGGTACGCACCAACAGACAGGCGCAAGATAATCAAGACAGACACAA	4219
Qy	3598	accgctgcgcagcagcttggtgatgtgcttaaacgcagggctttaactctgcaaggcaatgggt	3657
Db	4220	ACCGCTGCAGCAGCATGTTGATGTGCTTAAGCGCAGGGCTTTAACTTCAAGGCAATGGT	4279

Qy	3658	gaagcgttgactttgtctccacttatgacactgttgactttatogatggcaatgccacc	3711
Db	4280	GAAGCGGTTGACTTTGTCCTCACTTATGACACTGCAACTTTTTGGCATGGCAATGCCACC	4339
Qy	3718	accgctaagctgacctatgatgacacaagcaaaacccagtaaaagtggctatgatgtcaat	3777
Db	4340	ACCGCTAAGTGACCTATGATGNACACAAACAAACAGTAAGATGGTCTATGATGTCAAT	4399
Qy	3778	gtgataataaaaccattgaagtgcacaagtataaaaaacttggcgtcaaaaccaccaca	3837
Db	4400	GTGGATGATCAACCACTTGAAGTTAAA---GATAAAAAACTTGGCGTAAAAACCAACCACA	4456
Qy	3838	ctgacaaaaacaagtactaatggtaatgca---accaaatttagtgcgcgcgatggcgat	3894
Db	4457	TTGACCAAGTACTGGCAGAGGTGCTAATAAATTTGCCCTTAAGCAATCAAGACTTACTGGCGAT	4516
Qy	3895	gcccttgtaaagccagtgatatcgccacccatctaaataccttggctggcgacatccaa	3954
Db	4517	CGGCTTGTCGAAGCGCAGTGATCTGGTGTCTATCTTAACACCTTATCTHGGCAGCATCCAA	4576
Qy	3955	accgccaaaagggcgaagccaagcaagcagctcagcaagctatgtgatgtgatggcaac	4014
Db	4577	ACTGCCAAAGGGCAAGCCAAAGCAACGCTCAGCAGGCTATGTGGATGCTGATGGCAAT	4636
Qy	4015	aagtcactctatgacagtcaccgataaagaagtactatcaagtcoaatgacaagggctcaagt	4074
Db	4637	AAGTCACTATTCACAGTACCGGATAACAAGTACTATCAAGCCAAAAATGATGGCACAGTT	4696
Qy	4075	gacaaaaaacaagaagttgccaaagcaaaactggctgcgaagcccaaaccccaagatggc	4134
Db	4697	GATAAAAACCAAGAGTTGCCAAGAGCAAACTGGTGGCCCAAGCCCAAAACCCCAAGATGGC	4756
Qy	4135	acattggctcaaatgaatgtccaaatcagtcattacaagaagcagaagtaaatgagtccaat	4194
Db	4757	ACATTGGCTCAATGAATGTCAAAATCAGTCATTAAACAAGAACAAAGTAATGATGCCAAT	4816
Qy	4195	aaaagcaaggtcfaatgaagacaacgccttatacaaaagggcttgaaaaacgcgcgcaaa	4254
Db	4817	AAAAAGCAAGGCATCAATGAAGACAACGGCTTTGTTTAAAGGACTTTGAAAAAGCGCGTCT	4876
Qy	4255	gacacaaacccaaaaacgcgcagtaactgtgggtgatttaaatgcggttgccaaaca	4314
Db	4877	GATAACAAAAACCAAAACGGCGCAGTAAGTGTGGGTGATTTAAATGCCGTTGCCCAACA	4936
Qy	4315	ccgctgacctttgcagggatagcgcacaaacgcgtcaaaaaactggcgagactttgacc	4374
Db	4937	CCGCTGACCTTTGCAGGGGATACAGGCACACACGGCTAAAAAACTGGCGGACACTTGACC	4996
Qy	4375	atcaaaagtgggcaaacagacaccaataagctaaccgataataaacatcggtgtggtagca	4434
Db	4997	ATCAAAAGTGGGCAACAGACACCAATAAGCTAACCGGATAAATCAACATCGGTGTGTAGCA	5056
Qy	4435	ggtactgatggcttcaactgtcaaaacttgccaagaccctaaccactcttaacagcggttaat	4494
Db	5057	GGTACTGATGGCTCACTGTCAAACTTGCCAAAGACCTTAACCAATCTTTAACAGCGTTAAT	5116
Qy	4495	gcaggtggccaccgaattgatgaaaaggcatctctttgtagacgcgaacgcgtcaagcc	4554
Db	5117	CGAGCTGGTACCCAAAATTGATGACAAGGCGTGCTCTTTTGTAGACTCAAGCGGTCAAGCC	5176
Qy	4555	aaagcaaacacccctgtgctaaagtgccaatggcctggacactgggtggcgaacgcgatcagt	4614
Db	5177	AAAGCAAAACCCCTGTGCTTAAGTGCCAATGGCTGGACCTGGGTGGC-----	5224
Qy	4615	aacatcggtgcagctgtgatgataacgatcggtggaactttaagcaggttaatgaagtt	4674
Db	5225	-----	5224
Qy	4675	gccaaaacggtcaacaacctaaacaccaaagtaactcaggtgcgtcattaccctttgtg	4734
Db	5225	-----	5224
Qy	4735	qtaaccatgccaatggcaagcccataatggcaccagatggcaagcccaaaaagccatc	4794

Db	5225	-----	5224	5459	AAACTTGGCCACTGGTGGTGTACAAAGTGGCGGTGATTAAGACGGCAACGCTTAACGGCGAT	5518
Qy	4795	aagggcgccgatggttaataactatcacgccaaacggaacggtacgtgtggacaagaat	4854	5875	taaagcaatgtttgggtcaaaaccccaaaagatggcagcaaaaagccctgctcggcact	5934
Db	5225	-----	5224	5519	TTAAGCAATGTTTGGGTCAAAACCCCAAAAGATGGCGCAAAAGCCCTGCTCGCCACT	5578
Qy	4855	ggcaagcccatcacgcatgcggaacaaacttgccaatctggcagctcatggcaaacccctt	4914	5935	talaagcgccgaggtcagaccactatttgacaaacaccccgagagaagccattgacaga	5994
Db	5225	-----	5224	5579	TATAACGCCCGCAGGTTCAGACCAACTATTTCACCAACAACCCCGCAGAACCCATTGACAGA	5638
Qy	4915	gatgcaggtcatcaatggtggaagccctaggcggaactcaagatgacctcacccctaacc	4974	5995	ataatgaacaaggttatccgctctctccatgtcaacagatggcaatcaagagctggtga	6054
Db	5225	-----	5224	5639	ATAAATGAACAAGGTATCCGCTTCTTCCATGTCAACGATGGCAATCAAGAGCCTGTGGTA	5698
Qy	4975	aacatcaagtccactttgccaaaattgacacacacacacaggttaatgccaatgcagg	5034	6055	caagggcgtaacggcattgactcaagtgcctcagggcaagcactcagtgggcagatggttc	6114
Db	5225	-----	5224	5699	CAAGGGCGTTAACGGCATTTGACTCAAGTGCCTCAGGCAAGCACCTCAGTGGCGATAGGTTTC	5758
Qy	5035	caagcccaagctgcccagcctatcagcagcacagcaacacacaggttaatgccaatgcagg	5094	6115	caggccaaggcagatggtgaagccgcgttgccataggcagacacaaacccaagcagcaac	6174
Db	5225	-----	5224	5759	CAGCCCAAGGCAGATGTTGAAGCGCGCTTGGCATAGGCAGACAAACCCCAAGCAGCAAC	5818
Qy	5095	gagtgtcaaatgtaggtttaacttgcagaccacaaatcacaaatcaagtggactttgtcaaa	5154	6175	caatccatcgccatcggtgataacgcacaaagccacggcgatcaatccatcgccatcggt	6234
Db	5225	-----	5224	5819	CAATCCATCGCCCATCGGTGATAACGCACACAGCCACAGCGCATCAATCCATCCGCAATCGGT	5878
Qy	5155	gcctatgataccgtcaactttgtcaatggttacaggtgcccagatcaacagcgtgctgtagt	5214	6235	acaggcaatgtggttaacaggtgaagcactctggtgcctcggcaccgacccaagcactgtaag	6294
Db	5225	-----	5224	5879	ACGGCAATGTTGAGCAGGTAAAGCACTGTGTGCCATCGCGCACCCCAAGCAGCTTTAAG	5938
Qy	5215	gctgatggcagatgagtaacatcacgtcaacacccgcttagcagcgaccgatgatgat	5274	6295	gctgataacacttacgtgtggttaataacaacacagtttatcgatgcactcagaccgat	6354
Db	5225	-----	5224	5939	GCTGTATAACAGTTACAGTGTGGGTAAATAACAACCCAGTTTACCGATGCCACTCAACCCGAT	5998
Qy	5275	ggcaatglttataaagccaaagatggttaagtctacaagcagacagacctcatgcca	5334	6355	gtctttggtgtgggcaataacatcacogtgaccgaagaagtaactcgggttgcttaggttca	6414
Db	5225	-----	5224	5999	GTCTTTGTTGGTGGGCAATAACATCACCGTGACCGCAAAAGTAACCTGCTGCTTGGTTCA	6058
Qy	5335	aacggtcactaaaagcaggcaaatcagcagtgatgcacacccgcttagcagcctcaactcgaagc	5394	6415	aactgtccatcagtcagcacacacacagcgacacacagcgaacgaacaaatctgacggcaca	6474
Db	5225	-----	5224	6059	AACTCTGCCATCAGTGCAGGACACACCCAGCCACACAAAGCCCAAAATCTGACGGCACA	6118
Qy	5395	ctgttaaccccaatgctggttaaaggcagtcacagcgatgcagtggtctcttaactta	5454	6475	gcaggtacaacacacacagcaggtgcaacaggtacgggttaaaggcttgctgggcaaacg	6534
Db	5225	-----	5224	6119	GCAGGTACAAACCAACACACAGCAGGTGCAACCGGTACCGTTAAAGCGTTTGGTGGACAAACG	6178
Qy	5455	tcaaaagcggattttaataccaagaatggttacaactactaccacagtaagctctgtgagc	5514	6535	ggggtgtgtggtctcgtgtggtgtcctcaggtgctgaacgcgtatccaaaatgtggca	6594
Db	5225	-----	5224	6179	GGCGTGTGTGCGGTCTCCGTGGGTGCTCAGGTGCTGAACCGCGTATCCAAAATGTGGCA	6238
Qy	5515	atcagttaccaaggcaagaataacagcagcatcaccccttaagcaagaatgggctgaatga	5574	6595	gcaggtgaggtcagtgccaccagcaccggtcggtgcaatggtagcagttgttacaagacc	6654
Db	5225	-----	5224	6239	GCAGGTGAGGTGAGTGGCCACACAGCAGCGATCGGTCAATGGTGAAGCGTTGTACAAAGCC	6298
Qy	5575	ggcgttaaggttcacagaatggtggttaaaaggacacacacacacacacgctgccaatgta	5634	6655	acccaaggcatgccaacgcacccaatgagcttgaccatcgatccacacacacacacacacac	6714
Db	5225	-----	5224	6299	ACCCAAGGCAATGGCCAACCAACCAATGAGCTTGACCATCGTATCCACCAACCAACCAAT	6358
Qy	5635	caacagttaaacgaagtagcgaactgtgtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	5694	6715	aaagccaatgcaaggatttcacagcgtatggcgatggcggtccatgccacaagcctacatt	6774
Db	5279	CACAGTTAACGAGTAGCAGCAACTTGTGGTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT	5338	6359	AAAGCCANTGCAGGGATTTTCATCAGCGATGGCGATGGCGTCCATGCCACAGCCTACATT	6418
Qy	5695	gacggcaatcaggttaaacattgccagatcaacacacacacacacacacacacacacacac	5754	6775	cctggcagatccatggttaaccgggggtattggccacacacacacacacacacacacacacac	6834
Db	5339	GACGGCAATCAGTAACATTCGCCGACATCAAAAAGACCCCAAAATTCAGGTTCAATCAATCT	5398	6419	CTTGGCAGATCCATGGTTACCGGGGGTATTGGCCACCCACCAACCGTCAAGTCCGCTGGCA	6478
Qy	5755	aaccgacgtcatcaaaagcagcagcaggttacttggcggttaaaaggttaataacagataccga	5814	6895	gatacccaagcccatgtagggggcgaggtggtgaggggttttcacattt	6942
Db	5399	AACCGCAGTGTATCAACAGCAGCGCAGCGTACTTGGCGGTAAAGGTAATAACGATACCGAA	5458	6539	GATACCCAAGGCCATGTAGGGGCGCAGTTGGTGCAGGTTTTTTCACCTTT	6586
Qy	5815	aaacttgccactggt	5874			

[illegible]

[illegible]


```

ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5738 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-685-467-3

```

RESULT 10
US-08-478-370-4
Sequence 4, Application US/08478370
Patent No. 5808024
GENERAL INFORMATION:
APPLICANT: SASAKI, Ken
APPLICANT: HARKNESS, Robin E.
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
MEMBRANE PROTEIN OF MORAXELLA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario

RESULT 9
US-08-685-467-3
; Sequence 3, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States

; PRIOR APPLICATION NUMBER: GB 9726398.2
 ; PRIOR FILING DATE: 1997-12-12
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 1776
 ; TYPE: DNA
 ; ORGANISM: Neisseria meningitidis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1773)
 US-09-377-155-10

Query Match 0.8%; Score 55.2; DB 4; Length 1776;
 Best Local Similarity 51.2%; Pred. No. 7.6e-06;
 Matches 129; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY	6689	accatcgatccaccacaaacgaaataaaagccaaatgcagggtttcatcagcgatggcga	6748
DB	1520	acaacgcgcatcgacaatgtggacggcaacgcgctgaggcatcgcccaagcgattgcaa	1579
QY	6749	tggcgtccatgccacaagcctacattcctcggcagatccatggttacgggggtatgccca	6808
DB	1580	ccgcaggctcgttcaggcgatttgcgcggcaagagatgatggcgatcgcgcgcgca	1639
QY	6809	cccaacacggtcaaggtgcggtggcagtggtgactgtcgaagctgtcgggataatggtaaat	6868
DB	1640	cttaccgcgcggaagcgggttaacgccatcggtactccagatttccgacgcggaatt	1699
QY	6869	gggtatttaaaatcaatggttcagcgcgatacccaagcccatgtaggggcgcgagttggtg	6928
DB	1700	ggattatcaaaaggcagcgttcgcggcaattcgcggccatttcggtgtcttcgcgcatctg	1759
QY	6929	caggttttcaact	6940
DB	1760	tcggttatcagt	1771

Search completed: September 12, 2001, 13:00:35
 Job time: 1257 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2001, 03:34:48 ; Search time 18125.9 Seconds
(without alignments)
3620.328 Million cell updates/sec

Title: US-09-361-619-10
Perfect score: 6942
Sequence: 1 atgaatacacatctataaagt.....ttggtgcaggtttcaacttt 6942

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

Thu Sep 13 14:18:05 2001

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[illegible]

Query Match 0.8%; Score 55.6; DB 219; Length 1101;
Best Local Similarity 21.1%; Pred. NO. 0.0018;
Matches 115; Conservative 191; Mismatches 232; Indels 6; Gaps 2;

RESULT	6
AQ940248	
LOCUS	
DEFINITION	AQ940248 700 bp DNA GSS 27-JAN-2000 Sheared DNA-42E21.TF Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-42E21, DNA sequence.
ACCESSION	AQ940248

[illegible]


```

QY 1680 aattggcttgcgtgctgataagttgataaaagtcaccatatttggataaaaa 1739
Db 559 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 618
QY 1740 acaactcagtggtggtttaaataaccagagcagtgccatgaatcaggtgatca 1799
Db 619 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 678
QY 1800 aagatcagtaagtgttaaaat 1821
Db 679 TAATAATAATAATAATAATAATAA 700

RESULT 7
CNS0071A/c 895 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
fly); genomic survey sequence.
ACCESSION AL066286.1 GI:4945153
VERSION 1
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
  ORGANISM Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  REFERENCE 1 (bases 1 to 895)
    AUTHORS Direct Submission
    TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
    JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
    COMMENT - Web : www.genoscope.cns.fr)
      Determination of this BAC-end sequence was carried out as part of a
      collaboration with the Berkeley Drosophila Genome Project (BDGP).
      The BDGP is constructing a physical map of the Drosophila
      melanogaster genome using these BACs. For further information
      please see http://www.fruitfly.org The BDGP Drosophila
      melanogaster BAC library was prepared by Kazutoyo Osogawa and
      Aaron Mammossier in Pieter de Jong's laboratory in the Department of
      Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
      NY. The library is named RPCI-98 and was constructed by partial
      EcoRI digestion of Drosophila DNA provided by the BDGP from the
      isogenic strain y2; cn bw sp, the same strain used for the BDGP's
      p1 and EST libraries. A more detailed description of the library
      and how to order individual BAC clones, the entire library, or
      filters for hybridization from the BACPAC Resource Center can be
      found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
  source Location/Qualifiers
    1..895
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone_lib="RPCI-98"
      /clone="BACR14B09"
      /note="end : TET3"
BASE COUNT 124 a 80 c 204 g 179 t 308 others
ORIGIN
Query Match 0.8%; Score 54.2; DB 219; Length 895;
Best Local Similarity 19.2%; Pred. No. 0.004;
Matches 74; Conservative 139; Mismatches 172; Indels 0; Gaps 0;

QY 2380 accacccgtaaggttaacttacgtaaaccaatacaaccagtaagtaagtattgatgtc 2439
Db 859 MMMAACCCMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 800
QY 2440 aatgtggtgatgagaaaccattgaactcacagcgcgatgaatggcagacaacaaatggc 2499
Db 799 MCMCMACMMCMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 740
QY 2500 gtcaaaaccaccacactgaccacaacaaatgctaagtgttaagcaacacactttagtacc 2559
  1 : |||::: : : : | : | : | | | ||| : : :

```

```

Db 739 MCMCMCCCMAMMMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCM 680
QY 2560 accgataacgatgccttggtaaagcgaagacatcgccgaaatctaaacacccctagcc 2619
Db 679 CAMCMCMCCMACACMMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 620
QY 2620 aaggaattcacaccacaaagcagcagcagcagcagcagcagcagcagcagcagcagc 2679
Db 619 ACCMMCMCMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 560
QY 2680 aaagacggtgtaactgatgacgaacacatcacctggtggtggtggtggtggtggtggt 2739
Db 559 MCCAMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 500
QY 2740 aagacgtcaacactctaaactca 2764
Db 499 MMMAACCCMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 475

RESULT 8
AZ550599 884 bp DNA GSS 14-NOV-2000
LOCUS ENTGK07TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ550599
VERSION AZ550599.1 GI:11175900
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
  ORGANISM Entamoeba histolytica
    Eukaryota; Entamoebidae; Entamoeba.
  REFERENCE 1 (bases 1 to 884)
    AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
    TITLE Determination of clone end sequences from Entamoeba histolytica
    JOURNAL HMI:IMSS sheared DNA library
    COMMENT Unpublished (2000)
      Contact: Brendan J. Loftus
      Department of Eukaryotic Genomics
      The Institute for Genomic Research
      9712 Medical Center Dr., Rockville, MD 20850, USA
      Tel: 301 838 0200
      Fax: 301 838 0208
      Email: bjl@tigr.org
      Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
      DNA library
      Seq primer: M13-Reverse
      Class: shotgun
      High quality sequence start: 5
      High quality sequence stop: 860.
FEATURES
  source Location/Qualifiers
    1..884
      /organism="Entamoeba histolytica"
      /strain="HMI:IMSS"
      /db_xref="taxon:5759"
      /clone_lib="Entamoeba histolytica Sheared DNA"
      /note="Vector: pHOSt; Site 1: Bst I; Constructed at The
      Institute for Genomic Research (TIGR), Rockville, MD.
      Genomic DNA isolated from broth cultures of E. histolytica
      using a method described by Clark and Diamond (Clark,
      C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
      method for isolate identification. Exp. Parasitol.
      77:450.). The DNA was mechanically sheared to give a
      tight size distribution (~2 kb). The v + i method used for
      the library construction is described in detail in Smith,
      H.O. and Venter, J.C. (Making small insert libraries for
      whole genome shotgun sequencing projects. In Genome
      Sequencing: A Practical Approach, eds. M. Vaudin and B.
      Barrell, Oxford University Press, 1999)."
BASE COUNT 443 a 94 c 151 g 196 t
ORIGIN
Query Match 0.8%; Score 53.6; DB 245; Length 884;
Best Local Similarity 47.3%; Pred. No. 0.0058;

```


Matches 195; Conservative 0; Mismatches 214; Indels 3; Gaps 1;

QY 1446 aattggttttagcaataaagctggtacagttgatgaacacacaccccttatcttgataaaga 1505
 Db 289 AGTAGATGATGAATTTTAGAAGAAAAGAAAAGAAAATCAGCAGAGCTTGNATTAGA 348
 QY 1506 caagctaaagttgcaacagcaccctaaacacggtggtgctgactgttgaataacacccat 1565
 Db 349 AATATAAAAGATGATCTTGTGATTTACAAGCAATGATCAAAATCTTAAACAAAATCT 408
 QY 1566 tgggtggaagcaataaacaagctcgggtgctgagtggaattaaatttg---ccgatgt 1622
 Db 409 TCGTGGATTACACAGAAATTAATAAATAAAGAAAGAAATCATCACTGAACACAGATAA 468
 QY 1623 gaatgttaagtatcaaatgcccgaacaaattcgccactactcgtattaccgaagaggaat 1682
 Db 469 ATCTAATAAAGATTTAGCAGCTCAAAAAGTTTATCAAGAACAAAAGAAAATTTAGMAAG 528
 QY 1683 tggcctttgctgctgattggttaagttgataaagctcaccatatttggtataaacaaca 1742
 Db 529 TGAATTTGGCTGATGTTAAATTAATTAAGACACAACTCAACAAAGAACTGGTGTGCTACTCA 588
 QY 1743 acttcaagtggtggtttaaattaccacagcagtgccattatgcagggtgatacaaaa 1802
 Db 589 AGCTCGAGCTGATGGAATTAAGAAAGAAATTCAGGATATTACTCAGGAACAGATGGATG 648
 QY 1803 gatcagtaattgtaagatgcaacgacgatacccgatgcagtcactataaa 1854
 Db 649 GATTAGACAGGCTAAGAGCAAGTAAGCAAAATAGATTCAATTAAATGACTGAA 700

RESULT 9
 LOCUS AU088119 500 bp mRNA EST 27-JAN-2001
 DEFINITION AU088119 Sugano Malaria cDNA library Plasmodium falciparum cDNA clone XPfn6549, mRNA sequence.
 ACCESSION AU088119
 VERSION AU088119
 KEYWORDS AU088119
 SOURCE ESI.
 ORGANISM malarial parasite P. falciparum.
 REFERENCE 1 (bases 1 to 500)
 AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
 TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum
 JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
 MEDLINE 20574754
 COMMENT Contact: Junichi Watanabe
 Institute of Medical Science
 The University of Tokyo, Department of Parasitology
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Tel: 81-3-5449-5378
 Fax: 81-3-5449-5410
 Email: jwatanab@nagae.ims.u-tokyo.ac.jp
 Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES
 Source 1, 500
 /organism="Plasmodium falciparum"
 /strain="3D7"
 /db_xref="taxon:5833"
 /clone="XPfn6549"
 /dev_stage="erythrocytic stage"
 Location/Qualifiers
 214 a 37 c 86 g 163 t

BASE COUNT 214 a 37 c 86 g 163 t
 ORIGIN
 Query Match 0.8%; Score 53.4; DB 107; Length 500;
 Best Local Similarity 46.0%; Pred. NO. 0.0052;
 Matches 180; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 1459 aataaagctggtacagttgatgaacacacaccccttatcttgataaagcaagctaaaagtt 1518
 Db 38 AATGATGATGATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 97
 QY 1519 ggcacagcagcccttaacacacggtggtgctgactgttaataacacacccatcgggtgtagcaat 1578
 Db 98 AATGATAATGACAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 157
 QY 1579 aaacaaatccaaagtcggtgctgagtgccattaaatttgcgagtgatgaattgtaattgtaattca 1638
 Db 158 AATGATGATGACAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 217
 QY 1639 aatgcgcgaataatcgccactactcgtattaccgaagaggaataatggcgttctgctgactgt 1698
 Db 218 AATGATGATGACAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 277
 QY 1699 gatggttaagttgataaagtcaccatatttgataaacaacacacttcaagtggtggtggt 1758
 Db 278 GATTCCAAATCTTCATGCTAGTTCTTCTTAATATGCTTAATATTAATAACACGCGTGATAAT 337
 QY 1759 gttaaaattaccacagcagtgccattatgcagggtgatacaagatcgatgtaattgtaaa 1818
 Db 338 AATTCAGTAACTAACTTCTAGTAAATTAATAATACAGCAGCAAAATTAATAATGATGATGAT 397
 QY 1819 gatgcaacggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1849
 Db 398 ATTGTAGATCAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 428

RESULT 10
 LOCUS CNS00600/0 884 bp DNA GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL065923
 VERSION AL065923
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 884)
 AUTHORS Genoscope
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2; cn bw sp, the same strain used for the HDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 Source 1, 884
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR14N21"
 /note="end : T7"
 Location/Qualifiers
 230 a 62 c 139 g 124 t 329 others
 BASE COUNT 230 a 62 c 139 g 124 t 329 others

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:37:48 ; Search time 78.73 Seconds
(without alignments)
1781.835 Million cell updates/sec

Title: US-09-361-619-11
Perfect score: 11694
Sequence: 1 MNHKKVFNKATGTGTFMAVA.....NGSADTQGHVGAAGVGFHF 2314

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	11694	100.0	2314	22 AAB69136	M. catarrhalis les
2	7219.5	61.7	2053	22 AAB69135	M. catarrhalis str
3	6593.5	56.4	2047	22 AAB69134	M. catarrhalis str
4	6407.5	54.8	1952	22 AAB69133	M. catarrhalis str
5	6404.5	54.8	1992	17 AAW04505	Moraxella 200 kDa
6	6404.5	54.8	1992	22 AAB69137	M. catarrhalis M56
7	1177.5	10.1	2353	17 AAR99393	Haemophilus adhesi
8	1175.5	10.1	2411	21 AAB23860	Haemophilus influe
9	968	8.3	1104	21 AAB23856	Haemophilus influe
10	968	8.3	1104	21 AAB23859	Haemophilus influe
11	959.5	8.2	1004	21 AAB23857	Haemophilus influe

12	895.5	7.7	1002	21 AAB23854	Haemophilus influe
13	840.5	7.2	2039	19 AAW56322	Haemophilus paraga
14	796	6.8	2042	19 AAW56319	Haemophilus paraga
15	708	6.1	1094	21 AAB23858	Haemophilus influe
16	659.5	5.6	1098	17 AAR99392	Haemophilus adhesi
17	592	5.1	679	17 AAR99394	Haemophilus adhesi
18	592	5.1	679	21 AAB23855	Haemophilus influe
19	564.5	4.8	1601	18 AAW30292	Non-typeable Haemo
20	546	4.7	2514	21 AAW75097	Neisseria meningit
21	521	4.5	1598	18 AAW30291	Non-typeable Haemo
22	518	4.4	1529	14 AAR41732	High molecular wei
23	504.5	4.3	2599	21 AAY75098	Neisseria meningit
24	485	4.1	1536	18 AAW30293	Non-typeable Haemo
25	482	4.1	1536	14 AAR41723	High molecular wei
26	481	4.1	1536	14 AAR41725	High molecular wei
27	480	4.1	1536	15 AAR63505	Haemophilus high m
28	480	4.1	1536	21 AAB01846	Haemophilus influe
29	475	4.1	1477	14 AAR41724	High molecular wei
30	463.5	4.0	2383	21 AAB15945	E. coli proliferat
31	463	4.0	1477	15 AAR63506	Haemophilus high m
32	463	4.0	1477	18 AAW30294	Non-typeable Haemo
33	461	3.9	1477	14 AAR41728	High molecular wei
34	461	3.9	1477	21 AAB01848	Haemophilus influe
35	459.5	3.9	2893	19 AAW98828	H. pylori GHPO 148
36	459.5	3.9	2893	19 AAW71556	Helicobacter poly
37	458.5	3.9	1638	20 AAY00138	Enterococcus faeca
38	458.5	3.9	1638	20 AAY00140	Enterococcus faeca
39	458.5	3.9	1638	20 AAY00142	Enterococcus faeca
40	452.5	3.9	1338	14 AAR41731	High molecular wei
41	448	3.8	3596	21 AAY87407	Bordetella pertuss
42	444	3.8	3647	11 AAR05041	Filamentous haemag
43	434	3.7	1981	19 AAW42634	Protein sequence t
44	433.5	3.7	2902	22 AAB46351	H. pylori HPN165 p
45	432	3.7	1188	21 AAB01844	Haemophilus influe

ALIGNMENTS

RESULT 1
AAB69136
ID AAB69136 standard; Protein; 2314 AA.
XX AAB69136;
XX AC
XX 24-APR-2001 (first entry)
XX M. catarrhalis les1 200kDa protein SEQ ID NO:11.
DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.
KW Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
XX
XX WO200107619-A1.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-CA00870.
XX
XX 27-JUL-1999; 99US-0361619.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX WPI; 2001-159722/16.
XX N-PSDB; AAF59105.
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis

PS Claim 1; Fig 5A-Y; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. CC The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (II), and its truncated versions, are used as CC immunogenic compositions and vaccines to protect against M. catarrhalis CC infections, particularly otitis media in humans. (II) is also used as CC antigen in immunoassays for detecting specific antibodies (Ab), and to CC generate Ab. (I) are used for recombinant production of (II) and its CC fragments are used as probes for identifying/cloning 200 kDa protein. CC genes from other strains, and for diagnostic detection of M. catarrhalis. CC (I) makes possible production of large amount of recombinant immunogens. CC Expression of truncated versions of (II) reduces toxicity of the protein CC towards the Escherichia coli host. The present sequence represents the CC M. catarrhalis les1 200kDa protein, which is given in the exemplification CC of the present invention.

XX. Sequence 2314 AA:

Query Match 100.0%; Score 11694; DB 22; Length 2314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNHLYKVFENKATGTFMAVEACAKSHSGSSSTAGQVGSSPVIRLTRVATLAILVIGAT	60
Db	1	mmhlykvlfnkatgtfmaveacakshsgssstaggvsspvirltrvatlailvigat	60
Qy	61	LNGSAYAQNNKIAFGTTGNNDNASASNEASIAIGSLAKAHANOAIAGGSKPDRNQAA	120
Db	61	lngsayaqnnskiafgttgnndnasasneasiaigslakahanaqalaigskpdrnqea	120
Qy	121	NOKAGSHAKGKESIAIGDVLAEGDASIAIGSDDLXLDNRNSTNSKYPNGLLSTLIONHVV	180
Db	121	nqkeagshakgesiaigdvlaegdasiaigsdldlydrnstnksypngllstliqnhvt	180
Qy	181	LQIRDSNGSQKRYRTAEAGHASTAVGAMAYAKGHAFANAFGTRSTAEAGNYSIAVGLTAKA	240
Db	181	lrqirdsngsqkryrtaeaghasavgamayakghafanafgtrstaeagnysleavgitaka	240
Qy	241	EKGVTIAIGSNAQAIIYNGALAGDADTRVDLDYGTALGYGSGQILNNNNNNKAYVPEGNG	300
Db	241	ekgytiaigснааqiiyngalagadtrvldldygtalgygsgqilnnnnnnnkayvpegn	300
Qy	301	SNIKSKATGNGLFSTGSGTIRKRIINVGAGYEDTDVAVNAQLKAVENLAKRQITFKGDD	360
Db	301	snikskatnglflstgsgtirkrriinvgagyedtdavvnaqlkavenlakrqitfkqdd	360
Qy	361	NGTGVKKLGETITIKGGTQADKLTIDNNNIGVVDNNTGTLKVKLAKNLSGLETVSTKNL	420
Db	361	ngtgvkkigetitikggetqadkltidnnnigvvdnntgtlkvklaknlsgletvstcknl	420
Qy	421	TASEKTVVSGNNTAELOSGGLTFPTTNASTDVTYGTGDLKFTDNSNTALEDTTRITK	480
Db	421	tasektvvsgnntaelosggltfpttnastdvtvygtgdlkftdnsntaledttritk	480
Qy	481	DKICFSNKACTVDENKPYLDKDKLKYGNSTLNNNGGLTVNNTGGSNKQIQVAGDGIKAD	540
Db	481	dkicfsnkatvdenkpyldkdklkygnstlnnnggltvnntggsknlqvagadgikfad	540
Qy	541	VNVNVSNAKFGTRITEEIGFADADGKVDKSPYLDKKQLQVGVKTIKDSGINAGDQ	600
Db	541	vnnvnsnaakfgtriteeigfadaadgkvdkspylldkqlqvgvkiktidsginagdq	600
Qy	601	KISNVKATDQDVAITYKQLKQVQDADGALQSFSIRDEKGGQEFFISNYSNGTNPNTFE	660
Db	601	kisnvkatdtdavtykqlkqvqdadgalqsfsirdekqgqeffisnysngntpntfe	660
Qy	661	TITFAGENGISINDIAKGVKVCIDPINGLTTPKLTIVGSDKDGKQTOLVIEQVAGSNDTK	720
Db	661	tittfagengisindiakgvkvvcidpilinglttpkltivgsdkdgktqlvleqvagsndtk	720

Qy	721	NIIRGLSPTLPSITNAGVTRTEQNTITTSDEKSKAASIGDILNTGFNLKNNSNSVGFV	780
Db	721	niirglsptlpsitnagvtrtteqntittsdedkskaasigdilntgfnlknnsnsvgfv	780
Qy	781	STYNTVDFIDGNATTAQVYDETQTSKVYDVNVDEKTELTDGNGKTNKIGVKTWTLT	840
Db	781	styntvdfidgnattakvyydetnqtksvtydvnvdektdeltdgngktnkigvktwtlt	840
Qy	841	TYNANGKATNESTTDNDALVNAKDIAENLNTLAKEIHHTKGTADTALQTFKVKKDGATDD	900
Db	841	ttynangkatensttdndalvnaakdiaenlnltakeihtktgadtalqtfkvkkgdatdd	900
Qy	901	ETITVKGDTQNGKTVNTLKLKGENGLTVANNDGTQVTFGINTQSGLKAGDSTTLNDGL	960
Db	901	ettitvkgdtqngkvtntlklkgengltvankdgtvtfngintsglkagdsttlndgll	960
Qy	961	SKNPASNEQIQVAGDGVKFAKVDKGNSTGIDGTSRTIKDQIGFTGANGSLDTTKPHLT	1020
Db	961	sknpasneqiqvagdgvkfakvdkgnstgidgtsrtdkqigftgangsldttkphlt	1020
Qy	1021	KDKLVGEVEITNTGINAGGKKITNIQSGDITTONSNDVAVTGRVYDLKTELESKINSAK	1080
Db	1021	kdklvgeveitntginaggkkitniqsgditqnsndavtgrvrydlkteleskinsaak	1080
Qy	1081	TAONSLHEFSVADEQNHFTVSNPYSSYDTSKTSVDITFAGENGITTKVNGKGVVRGIDQ	1140
Db	1081	tagnslhefsvadeqnhftvsnpyssydtksdsvditfagengittkvngkvvrvgidq	1140
Qy	1141	TGGLTTPKLVGNNGKGIIVIDSKQONTITGLSNTLANVNDGAGHALSQGLANDTOKT	1200
Db	1141	tgglttpklvgnngkglividskqngntitglsantlanvndgaghalssqglandtkt	1200
Qy	1201	RAASIGDVLNAGFNJQNGEAVDFVSTYDFIDGNATTAQVYDETQTSKVYDVNV	1260
Db	1201	raasigdvlnagfnlqngaeavdfvscydydfidgnattakvyydetqtskvvydvnnv	1260
Qy	1261	DNKTIEVTSDDKLGKVTITTKTSANGNATKFSAADGDALVKASDSTATHLNTLAGDQTA	1320
Db	1261	dnktievtssdklgvtttktsangnatkfsaadgdalvksadstathlntlagdqtca	1320
Qy	1321	KGASQASSASVYDADGNKVLYDSFDKKYQVNDKGQVNDKNEVAKDKLVAQAQTPDGT	1380
Db	1321	kgasqassasvydadgnkvlydsfdkkyqvndkgqvndknevakdklvaqaqtpdgtl	1380
Qy	1381	AQMNYSKVINKEQVNDANKKQGINEDNAFTKGLNNAKDTTKNAAVTVGDLNVAQTP	1440
Db	1381	aqmnyskvinkeqvndankkqginednafkglennaakdttknaavtvgolnvaqtp	1440
Qy	1441	TPAGDTGTTAKLGETLTIKGGOTDNTKLTNNIGVWAGTDGFTVKLAKDLTNLSYNAG	1500
Db	1441	tfagdtgtttaklgetltikggotdntkltndnigvvgtdgftvklakdltlnsynag	1500
Qy	1501	GTRIDEKGISFVDANGQAKANTPVLSANGLDLGKRRISNIGAAVDDNDVAVKQFNEVAK	1560
Db	1501	gtridekgisfvdangqakanpvlisangldlgkrrisnigaaavddndvavkfnevak	1560
Qy	1561	TVNNLNQNSGASLPFVVTDANGKPIINGTDGKPKOKAIKGADGKYHANANGVPVDDGK	1620
Db	1561	tvnnlnqnsngaslpfvvtadangkpiingtdgkpkokaikgadgkyhanangvpvddgk	1620
Qy	1621	PITDADKLANLAHAKPLDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPNTGNANAG	1680
Db	1621	pidadkylanlahakpldaghqvvvaslggnsdailtlnikstlpqidtpntgnanaga	1680
Qy	1681	QSLPSLSAAQOSNAASVKDVLNVGNFLOTNHNQVDVSKAYDTVNFVNGTGADITSVRSD	1740
Db	1681	qslpslsaagqsnaasvkdvlnvgnfLOTNHNQVDVSKAYDTVNFVNGTGADITSVRSD	1740
Qy	1741	GTMSNITVNTALAAATDDGNNVLIKAKDGKPYKADDLMPNGSLKAGSADAKTPGCLSLV	1800
Db	1741	gtmsnitvntalaatdddgnvlikakdgkpykaddlmpngslkagsadaktppgclslv	1800
Qy	1801	NPNAKGSTGDVALNMLSKAVFKSKDGTITTTTSSDGTISIQGKDNSSITLSKDLGNVGG	1860

Db 1801 nbnagksgtdavalmniskavfkskdgtttttvssdglisqgkdnssllcskdglnv9g 1860
QY 1861 KVISNMGKGTDRDAANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSNR 1920
Db 1861 kvisnvgkgktdaavnvqlnevrnllglnagnadnagnqvniadiikkdpnsgssnr 1920
QY 1921 TVIKAGTVLGGKNNDEKATLGGVGVGDKNANGDLNNVWVKQDKGSKKALLATYN 1980
Db 1921 tvikagtvlggkgnndteklatgvgvgvdkdngnangdlnnvwvktqdkgskkallatyn 1980
QY 1981 AAGTNTVLTNPAEADIRINEOGIRFPHVNDGNOEPVVOGRNGIDSSASGKHSVAIGFOA 2040
Db 1981 aagtcnyltunpaeadirineogirfphvndgnoepvvoqgrngldssasgkhsvaigfqa 2040
QY 2041 KADGEAAVAICROTQAGNQSIAIGNAQATCDOSIAIGTGNVVTGKHSGAIGDPSTVKAD 2100
Db 2041 kadgeaavaiqrtaqgnqsaigldnaqatgdsiaigtgnvvtgkhsaigdpstvkad 2100
QY 2101 NSYSVGNNOFIDATQDVFVGNNITVTSNSVALGSNAISAGTHIAGTQAKSGDTAG 2160
Db 2101 nsysvgnnfdiatqtdvfgvgnnitvtsnsvalgsnaisagthiagtgakskdgtag 2160
QY 2161 TTTTAGATGVKGFAGOTAYGVSVGASGAERRIONVAAGEVSATSDAVNGSOLYKATQ 2220
Db 2161 ttttagatgtvkgfagotaygvsvgasgaerrionvvaagevsatscdavngsqlykatq 2220
QY 2221 GIANATNELDHRIHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALG 2280
Db 2221 gianatneldhrihqnenkanagissamamasmpqayipgrsmvtggiathngggavag 2280
QY 2281 LSKLSNDGQWFKINGSADTQGHVGAAGVGFHF 2314
Db 2281 lsklsndgqwfkingadtqghvgaavagfghf 2314

RESULT 2

AAB69135 ID AAB69135 standard; Protein; 2053 AA.
XX AC AAB69135;
XX DT 24-APR-2001 (first entry)
XX DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.
KW Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX OS Moraxella catarrhalis.
XX PN WO200107619-A1.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-CA00870.
XX PR 27-JUL-1999; 99US-0361619.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
XX WPI; 2001-159722/16.
XX DR N-PSDB; AAF59104.
XX PS New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
XX Claim 1; Fig 4A-V; 247pp; English.
XX The present invention describes an isolated and purified nucleic acid (1)

CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain Q8 200kDa protein, which is given in the
CC exemplification of the present invention.
XX
SQ Sequence 2053 AA;
Query Match 61.7%; Score 7219.5; DB 22; Length 2053;
Best Local Similarity 64.1%; Pred. No. 0;
Matches 1553; Conservative 140; Mismatches 250; Indels 481; Gaps 38;
QY 1 MNHIYKVFENKATGTFMAVAPCAKSHSGSSSTAGOVSSPVIRLTVATLAILVIGAT 60
Db 1 mnhiykvfhnkatgctfmavaeyakshstgsgscatgvgsvrtisfariaalavivigat 60
QY 61 LNSAYAQN-NSKTAFTCTTGNNDN-----ASASNEASTAIGSLAKAHANOIAIGGSKPD 114
Db 61 lngsayeqittkieigtunkinnlkgdlatgeaslatgfskagsgaiaigskvkd 120
QY 115 PRQANQKAGSHAKGKSTAGIDGDLAEGDASIAIGSDLLYLDRN-STNSKYNGLLST 173
Db 121 p-nngsngnvgshakgnesiaigdvlaegdasiaigsdldlylpknldlknef----hk 174
QY 174 LIONHTVLRIOIRDS-NGSQYRRTAAEGHASTAVGANAYAKHEANAFGRSTAEKNSYL 232
Db 175 lihghelliklqstdgklyrrtraggshastavvamsyaqghfsnagfatgataeaysl 234
QY 233 AVGLTAKAEKGYTTAIGNSAQAINYGALALCADTRVDLDYGAIGYGSQILNNNNNNNK 292
Db 235 avglaaqatkssiaavgsnakanaafaataigtgnvtvnlgrvalgfgsqilrdndtas 294
QY 293 AYVPEGNSNIKSSKATGNG-----LFSIG-----SSTIKRIINVGAGYEDTDVAVNAQL 343
Db 295 ayvplgk-tladgkatrqgdstdfisgnnnnnssirrkilnvgagsrdtdavnaql 353
QY 344 KAVENLAKROITFK-GDDNGTGKVKKLGLGFTLTIKGGETQADKLTDDNNNGVWTDNNTGL 401
Db 354 klveelanrktfkdgddnnnsnsvverglngtcltik-gdaqtnalte-anigvvcn-gl 410
QY 402 KVKLAKNLSGLETVSTKNTLTASERKTVV-GSGNNTAEIQQSGGLTFTPTTNASTDKTVYGTD 460
Db 411 kvklakeltgitsvs-----atnkltvsnntnnnaelqsggltsptgtktdktyysid 465
QY 461 GLKFTDMSNT-ALEDTRITKDKIGFNSKAGTVDENKPYLDKDKLVGNSTLNGLTGVN 519
Db 466 glkftdmsnsiatkgtrickkigfagtdgvdskpyldneklkvgnstlsgsltn 525
QY 520 NFIGGSNKQIQVGADGKFKADVNVNVSN-AAKFTGTRTTEEEIGFADAGDGKSPYLD 578
Db 526 ntfg--nkqigvgangkfatvannvantsatvgtariteekigfagtdgvdqeqayld 583
QY 579 KKIQGVGVKITKDSGINAGDQKISNVKDATDPTDAVYTKLK----- 621
Db 584 kerikvrveittdsginaghnkigtgtngiantdavrtdavtikgldakptlnagdgisn 643
QY 622 ----- 623
Db 644 ngdlvdsgnnttptynlsvkttklnsgtsgnnkfsvsnahdnnsivtakdladylnkv 703
QY 624 QODADGALQSFISIR--DEKGOEFTISNLYSNGNTPTNFETITFAGENGISINDIAGKV 681
Db 704 netadsalpsfkvgndgnsnaltvgk-dtng---kftntklkgngvntnatrgt 759

QY	682	KVICIDPGLTPKLTGSDRGKLTQVIEQV---ASGNDTKNIIRGLSPITLPSITNAGGV	739
Db	760	tfigdsgnglttptkltvgdtnng-nrlviedqpsadgnstknikiiglsptlpslaspsg-	817
QY	740	RTEGNTTTSDESKSAASIGDILITGFLNKNNSNGFVSTVNTVDFIDGNATTAKVT	799
Db	818	rniaignti-eeekdksnaasddvlnagfnknngkdkdfvstydytdvdfidgnattatvt	876
QY	800	YDETNTQSKVTYDVNVDEKTIETLTGDNKTKNKIGVKYTTTLTTTNANAKATNFSTTDNAL	859
Db	877	ydeangtskvaydvnnvdektietltdngng-kqlvgvktikltetstngnattfstddhdal	935
QY	860	VNAKDIAENLTLAKELIHTTKTAOTAILOTQFKVKK----DGATDDETITVKGDKGTQNGKT	915
Db	936	vkasdiagnlntlaeeihctkgtantatqtfvkvkvdendkaddtnaitykdkgt-sgk-	993
QY	916	VNTLKLKGNGLTAVTNKDGTVTFGINTQSLKAGDSTTLNKGLSIKNPASNEQIQVGA	975
Db	994	vntlkkgngldktdkgtvtfgintqsgikagdstllnnglsikntasneqigvga	1053
QY	976	DGVKFAKVDKNGSSTGIDGTSRTIKDOIGFTGANGSLDTTKPHLTKDKLKVGEVEITNTG	1035
Db	1054	dgvkfawnnvgvagiidgttrtdreigftgungsidkskphlskd-----g	1101
QY	1036	INAGGKKTINIQSGDITQNSDNDVATGGRVYDLKTELESKINSAAKTAQNSLHEFSYADEQ	1095
Db	1102	inaggkktitniqsgelaknshdadvtkglydlktelenkissaktagnslhefsadeq	1161
QY	1096	CNHFTVSNPSSYDTSKTSVDITFAGENGITTKVNGVVRVGIDQTKGLTTPKLTVGNNN	1155
Db	1162	gnftvsnpyssydstsktsdvitfagengittkvngvvrvgidqtkgltpkltvgnnn	1221
QY	1156	GKIVIDSKGQNTITGLSNTLANVND-GAGHALSG-LANDTDKTRAASIGDVLNAGF	1213
Db	1222	kgivinsqngntitglsntlanvndkgsvrttegnniikdedktraasivdvlsagf	1281
QY	1214	NIQGNQAVFVSFYDVFIDGNATTAKVTYDPTSKVWYDVNVNKNKTIEVTSDDKL	1273
Db	1282	nlqngneavdvstvydvnfanngnttckvtyddtsktskvvydvnnvddtiev-kdkkl	1340
QY	1274	GKVTTLTKTSANGNATKF---SAADGALVKASDIATHLNTAGDTQAKGASQASSA	1330
Db	1341	gvkttltstgtgan--kfaisngatgdalykvasdivahlnitsgdiqtakgasqannsa	1398
QY	1331	SYVDADGNKVIYDSTDKKXYQVNDKGOVDKNKEVAKDKLVAQAQTPDGTLAQMNKVSIN	1390
Db	1399	gyvdadgnkviydstonkyqvakngdgtvdktkevakdklvagactpdtlaqmnvksin	1458
QY	1391	KEQVNDANKKQGINEDNAFIKGLNAAKDKTKNAAVTVGDLNAVAQTPLTFAGDTGTTA	1450
Db	1459	keqvndankkqginednafvkgleaaasdnktnaaavtvgdlnavaqtpltfagdtgltta	1518
QY	1451	KKIGEFILITKGGQTDNKLNTDNINIGVAGTCDGFTVKLAKDLNLSNVNAGTTRIDEKGIS	1510
Db	1519	kkigetltikggqtdcnkldtnnigvvgcdgftvklakdlntlnsvnagtkkidekgis	1578
QY	1511	FVDANGQAKANTPVLISANGLIDLGKGRISNIGAAVDDNDVNEKQFNEVAKVTNNLNQSN	1570
Db	1579	fvdangqakantpvlisangldl-----	1600
QY	1571	SGASLPFVVTDANGKPTINGTDGKPKQAIKAGDGKYYHANANGVPVDPDKGKPI	1630
Db	1601	-----	1600
QY	1631	LAAGKPLDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPNTGNANAGQAQSLPSLSAAQ	1690
Db	1601	-----	1600
QY	1691	QSNAAKSVKDLNVGFENLQTNHINQVDFYKAYDTVNEVNGTGADITSVRSADGTMSTIVNT	1750
Db	1601	-----	1600
QY	1751	ALAAATDDGCVNLIIKAKDGKFYKADDLMPNGSLKAGKSASDAKTPTGLSLVNPNAKGSTG	1810
Db	1601	-----	1600
QY	1811	DAVALNNLSKAVFKSKDGTITTTVSSDGIISIQKDNSSITLSKDLNVGCKVLSNVGKGT	1870
Db	1601	-----gqkvlsnvvgkgt	1612
QY	1871	KUTDAANVOQLNEVRNLLGLGNAGDNADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLG	1930
Db	1613	ktdaanvqqlnevrnllgl----gndnadgnqvniadiikkdpnsgsssnrtvikagtvlg	1669
QY	1931	GKGNNDTEKLATGGVQGVYDKDGNANGDLSNVWTKQDKGSKALLATYNAAGOTNVLTN	1990
Db	1670	gkgnndteklatggvqgvvdkdgnangdlnsvwtkdkgskallatynaaagotnvytn	1729
QY	1991	NPAAETDRINEGIRFFHVNDGNQEPVQVGRNGIDSSASGKHSVAIGFQAKADGAAVAI	2050
Db	1730	npaeadrineggrirffhvndgnqepvqvgvrngidssasgkhsvaigfqakadgaaava	1789
QY	2051	GROTAQGNOSIAIGDQNAQATGQDQSTAIAGTGNVVTGKHSGAIGDPPSTVKADNSYSVGNNO	2110
Db	1790	grtqagngqsiaigdnatqdgdsiaigtgnvvagkhsaigdpstvkdadnsysvgnnnq	1849
QY	2111	FIDATQDFVFCVGNNTVTFESNSVALGSNSAISAGTHACTQAKKSDGTAGTTTACATGT	2170
Db	1850	ftdatqtdvfgvgnntvttesnsvalgsnsaisagthagtqakksdgtagtttttagatgt	1909
QY	2171	VKGAGOTAVGAVSVGASGAERRIQNVAAGEVSATSTDAVNGSOLYKATQGTANATNELD	2230
Db	1910	vkqagotavgvsvgsgaerriqnvaaagevsatstdavngsqlykatqgsianaeneld	1969
QY	2231	HRIHQENKANAGLSSAMAMASMPQAVIPGRSMVGTGGIATHNGQGAVALGSLKSLSDNGQW	2290
Db	1970	hrihqenkanaglsamamampqayipgrsmvtggiaathngqgavavglslksdngqw	2029
QY	2291	VFKINGSADTQCHVGAAGVAGGFHF	2314
Db	2030	vfkingsadtqghvgaavgagfhf	2053
RESULT 3			
AAB69134 standard; Protein; 2047 AA.			
XX	AAB69134;		
AC	AAB69134;		
XX	24-APR-2001 (first entry)		
DT	M. catarrhalis strain 4223 genomic 200kda protein SEQ ID NO:7.		
XX	Moraxella catarrhalis.		
XX	Moraxella catarrhalis.		
OS	Moraxella catarrhalis.		
XX	WO200107619-A1.		
PN	01-FEB-2001.		
XX	26-JUL-2000; 2000WO-CA00870.		
PF	27-JUL-1999; 99US-0361619.		
XX	(CONN-) CONNAUGHT LAB LTD.		
XX	Loosmore SM, Sasaki K, Yang Y, Klein MH;		
XX	WPI; 2001-159722/16.		
DR	N-PSDB; AAF59102, AAF59103.		
XX	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,		
PT			


```

PT useful in protective vaccines and for diagnosis -
XX Claim 1; Fig 3A-W; 247pp; English.
XX The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein (II) has antibacterial activity and
CC the 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain 4223 genomic 200kDa protein, which is given in the
CC exemplification of the present invention.
XX Sequence 2047 AA;
SQ
Query Match 56.4%; Score 6593.5; DB 22; Length 2047;
Best Local Similarity 60.7%; Pred. No. 2.6e-283;
Matches 1466; Conservative 149; Mismatches 330; Indels 471; Gaps 48;
QY 1 MNHIYKVFINKATGTFMAVEACSAKSHSGSSGTAGQGVSPVIRLTRVATLAILVIGAT 60
DB 1 mnhiykvfinkatgtfmaevaekshstggscatqgvsvctlsfariaalavlvgat 60
QY 61 LNSGSAQAQ---NNSKIAFGTTG---NNDNASASNEASTAIGSLAKAHANQAIAIGGSKPDP 115
DB 1 lgsayaqkdkthialgeqnprrgstakadgdraiaigenanagggaiaigssnktv 120
QY 116 RNQAANQKAGSHAKGRESIAIGDVLAEAGDASTAIGSDLLYLDORNSTNSKYPNG-LLSTL 174
DB 121 -ngsslckigtatggesiaigdvkaskgdaslaigsdldhldqgnphkpgxtlindl 179
QY 175 IONHTVLRIQIRDSNGSO-KYRRRTAAEGHASTAVCAMAYAKHFANFGRSTRSEAGNYSLA 233
DB 180 inghavleirsekdhndvkyrrittasghastavamsyaqghsfnaftgratakaysyla 239
QY 234 VGLTAKAEKYTTAIGSNAQAINYGALAGDTRVLDYGIALGYSGSIIUNNNNNNNKA 293
DB 240 vglataeagdstaigsdatssslgaialagagtraqlggsialgqgs-vvtgsdnnsrpa 298
QY 294 YVPEGNSNIKSKATCN--GLFSGISSTIKRKINVGAGYEDTDAVNVAOLKAVENLA 350
DB 299 ytpntqalpbk-fqatnntkagpplsigsnsikrkinvgagvknktadvnvaqlaevvkw 357
QY 351 K-RQITPKGDNDGTGVKKKLGSETITIKGG-ETQADKLTDDNNNIGVVD-NTTGLKVKLAK 407
DB 358 kerriitfggdndstvdkgidntltikggaetna-1td-nnlgvkeadnsglkvklak 414
QY 408 NLSGLETVSKNLTASKEVTVGSGNN-TAELOSGGLTFT-PTTNA-STDKTVYGTGLKF 464
DB 415 tlnlntevttlnattvtvkvgssstaelldslstftqptngsgstktvvygvngvkf 474
QY 465 TDNS-NTALEDTTRITKDKIGFSNKAGTVDENK-PYLDKDKLVGNSLTNNGSLTVNNTI 522
DB 475 tnaaetcaaigttrtrdklqga-rdgvdvdekapyldkklkvgsv-----altidngl 528
QY 523 GGSNKQITQVGADGKIFAD-VNVNVNNAKFCCTTRITBEEIGFADADGKVBKPSYLDKKQ 581
DB 529 dagnkklisnlakgssandavtieqlkaak--ptlnagagisvtpsteisvdaksgnvtapt 586
QY 582 LQGVGVKTKDSGINAGDQKISNVKDATDDTDVATYKQ----LKQVQDADGALQSFISIR 637
DB 587 ynl-gvkteelnsdgtstk--fsvkgsgttnslvtahlasylnevnrtadsalqstfvk 643
QY 638 DEKGOETISNLSNGTNP--TFETITFAGENGISISNDIAKGVKVGIDPDTNGLTTPK 695
DB 644 eedddanaity-akdttknagavsilklkgngltvatk-kdgtvtfglsqdsqglitgk 701
QY 696 LTVGSD---KDGKTLQVIVQVAVSGNDTKNI----- 722
DB 702 stlnndglvtkdnteq-igvgangikftvnvngsnptgiantarirtdrklgfsagdgav 759
QY 723 -----IRGLSPTLPSITNAGVTRTTEQGNITTSDE 752
DB 760 dtnkpyldqklvgvnykintntginaggkaigtlsptlsiadqss-rnielgniti-qdk 817
QY 753 DSKAASIGDITLANTGNFKNNSNSGVSVNTVDFIDGNATTAKYVYDETNOFTSKVTYD 812
DB 818 dksnaasindlntcfnlkanmnpdiftvdyldvfangnattatvtdtbanktskvvyd 877
QY 813 VNVDEKTIETLG-DNGKTNKIGVKVTTTTLTTTNANGK-ATNP--STTNDALVNKAIDAE 868
DB 878 vnvddttihigtddnk-klgvkttklntksangntatnfnvnssdedalvnakdaen 935
QY 869 LNTLAKEIHHTKTGADPALQTFVKK---DGATDDETTITVGKDGTOGKTVNTLKLKGE 924
DB 936 lntlakeihtktgtadtalgtftvkvdennnadaaitvggknann--qvntltlke 993
QY 925 NGLTVATNKDGTVTFTGINTOSGLKAGDSTTLNKDGLSTKNPASNEOIQVGADGVKFAKV- 983
DB 994 nglniktdkngtvtfgintstgikagks-tlndgglsiknptgseqiqvgadgkfakvn 1052
QY 984 DKGNSSTGIDTSRITKDQIGFTGANGSLDTTKPHLTKDKLVGEVEITNTGINAGGKKI 1043
DB 1053 nngvvgagidgttritrdeigftgtnsgldkskphlskd-----ginagggkkl 1100
QY 1044 TNSQSGDITQNSDNDVTVGGVRYDULKTELESKINSAAKTAQNSLHEFSVADPQGNHFTVS 1103
DB 1101 tniqsgeiaqnsdhavtggkiydlktelenkisstaktagnslhefsvadeqgnftvsn 1160
QY 1104 PYSSYDFTSKSDVITTFAGENGITTKVNGVVRVIGIDTQKGLTTPKLTGVNNGNGIVIDS 1163
DB 1161 pyssydcsksdvltfagengitctkvngvrvvgldqkglttkplkvgnngkgivids 1220
QY 1164 KDGNTTITGLSNTLIANYTND-GAGHALSQG-LANDTOKTRAASIGDVLNAGFNILQNGEA 1221
DB 1221 qngqntitglsntlianytndkgsrvrtteqgnlkdedktraasivdvlsagfnlqngnea 1280
QY 1222 VDFVSTYDVTDFIDGNATTAKYVYDFTSKTSKVYVYDVNVDNKTITEVTSDDKLGKVTTLT 1281
DB 1281 vdfvstydtvdfidgnattaktvdydtsktskvvydvndvddttiev-kdkklvgkttlt 1339
QY 1282 KTSANGNATKF---SAADGDALVKASDIATHLNTLAGDIOAKCASOASSASVVDADGN 1338
DB 1340 stgtgan--kfalsnqatgdalvksadlvahlnctisgdlqtakgasannsgvvdadgn 1397
QY 1339 KVIYDSTDKYYQVNDKGOVDKKNKEVARDKLVAQAQTPDGTLAQMNVKSVINKBQVNDAN 1398
DB 1398 kvlydstcnkyqakndgtvdktevakdklvaqaqtpdgtlaqmnyksvinkneqndan 1457
QY 1399 KKGINDENAFIKLENAADTKTKNAAVTVGDLNVAQPTLTTFAGDTGTTTAKKLGTEFLT 1458
DB 1458 kkgginednafvkllekaasdnktnaavtvvgdlvnaqptlttfagdtgtttakklgteflt 1517
QY 1459 IKGQOTPNKLTDDNIGVAVAGTDCFTVKLAKDLTNLSNVNAGGTRIDEBKGLSFVDANGQA 1518
DB 1518 ikggotpnkltddnigvavagtdcftvklakdltlnlsnvaggttklddkgvsfvdssgqa 1577
QY 1519 KANTPVLISANGDLGGKRISNIGAAVDDNDVAVNFKQFNEVAKTVNNLNQNSGASLPFV 1578
DB 1578 kantpvlisangldl----- 1591
QY 1579 VTDANGKPIGTGKPOKAIGADGKYHANANVPVDKDGKPIITDADKLANLAHAKPL 1638
DB 1592 ----- 1591
QY 1639 DAGHOVVASLGGNSDAITLTKSTLPLQIDPTTGNANAGQAQSLPSLSAAQQAQNAASVK 1698
DB 1592 ----- 1591

```


Thu Sep 13 14:18:07 2001

```

QY 1699 DVLNVGFNLQNHNOVFEVKAYDTVNFVNGTCADITSVRSADGTMSNITVNTALAAATDD 1758
Db 1592 ----- 1591
QY 1759 GNVLLKAKDKRFYKADLLMPNGLKAGKSASDAKTP7GLSLVNPNGKSTGDAVALNLL 1818
Db 1592 ----- 1591
QY 1819 SKAVFKSKDGTITTTVSSDGISIOGKDNSSITLSKDLGNVGKVIYNVKGTKDTDAANV 1878
Db 1592 ----- 1591
QY 1879 QQLNEVRNLLGLGNAGDNADGNQVNIADIKDPNSGSSNRTVIKAGTVLGGKGNNDTE 1938
Db 1612 qqlnevrnllglnagnadnagnvniadiikdpnsgssnrtvikagtvlggkgnndte 1671
QY 1939 KLATGCVQGVVDKGNANGDLSNVVVKQKDGSKKALLATYNAAGQTNVLTNNPAEATDR 1998
Db 1672 klatgqigvvdkgdngangdlsnvvvktgdkgskkallatynaagqtnvltnpaeaidr 1731
QY 1999 INEQGRFPHVNDGNOEPVVOGRNGIDSSASGKHSVAIGFQAKADGEAAVAIGROTQAGN 2058
Db 1732 ineqgrirfphvndgnoepvvogrnngidssasgkhsvaigfqakadgeaavaigrtqag 1791
QY 2059 QSTAIQDNQAQATGDSIAICTGNVVTGKHSGAIGDPSSTVKADNSYSVGNNGQFIDATOTD 2118
Db 1792 qstaiqdnqaatqdsiaigtgnvvtgkhsgaigdpstvkadnsysvgnngqfidatqtd 1851
QY 2119 VFGVGNITVTESNVALGSNSAISAGTAGTQAKKSDGTAGTITTTAGATGTVKGFAGQT 2178
Db 1852 vfgvgnitvtesnvalgsnsaisagthagtqakksdgtagttttagatgcvkfgagqt 1911
QY 2179 AVGAVSVGASGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQCIANATNHELDRHIHQEN 2238
Db 1912 avgavsvgasgaerriqnvaagevsatstdavngsolykatqcianatneldhrhihqen 1971
QY 2239 KANAGISAMAMAMPQAYIPGRSMVTGGIATHNGGAVAVGLSKLSDNGQWVFKINGSA 2298
Db 1972 kanagisamamampqayipgrsmvtggiathnggavavglsklsgdqgwvfkingsa 2031
QY 2299 DTQGHVGAAGVAGGFHF 2314
Db 2032 dtqghvgaavagghf 2047

RESULT 4
AAB69133
ID AAB69133 standard; Protein; 1992 AA.
XX
AC AAB69133;
XX
XX
DT 24-APR-2001 (first entry)
XX
DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kda protein SEQ ID NO:3.
XX
KW Moraxella catarrhalis strain 4223; major outer membrane protein;
KW 200kda outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX
OS Moraxella catarrhalis.
XX
PN W0200107619-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-CA00870.
XX
PR 27-JUL-1999; 99US-0361619.
XX
PA (CONN-) CONNAUGHT LAB LTD..
XX
PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
```

```

XX WPI; 2001-159722/16.
DR N-PSDB; AAF59100, AAF59101.
XX
PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX useful in protective vaccines and for diagnosis
PS Example 3; Fig 2A-W; 247pp; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kda outer membrane protein of Moraxella catarrhalis.
CC The 200 kda outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kda protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kda protein, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 1992 AA:
Query Match 54.8%; Score 6407.5; DB 22; Length 1992;
Best Local Similarity 60.5%; Pred. No. 4.2e-275;
Matches 1428; Conservative 146; Mismatches 316; Indels 471; Gaps 48;
QY 56 VIGATLNGSAYAQ--NNSKTAFTGTG---NNDNASASNEASIAIGSLAKAHANQAIAGG 110
Db 1 vigaatlsgsayaqkdktkhaigeqnqprsrsgtakadgdraiaigenanagaqgqalaigs 60
QY 111 SKDPRNOAANKAGSHAKGKESIAIGDVLAEADGASIAIGSDLYLDRNSTSKYKPNG- 169
Db 61 snktv-ngssldkigtatqgesiaigdvkagsdasiaigsdhlldghgnpkhpgkt 119
QY 170 LSTLIQNHVTVLRIQDSNGSQ-KYRRTAEGHASTAVGAMAYAKHFANAFGRSTAE 228
Db 120 lndlinghavlketrskdndvkryrrttasghastavgamsyqgghfnafgrataks 179
QY 229 NYSLAVGLTAKAEKGYTIAIGSNAQAINYGALAGADTRVDLYGIALGYSGSLNNNN 288
Db 180 ayslavglataegstiaigsdatssalgalaigagtraqlqgsialqggs-vvtqsdn 238
QY 289 NNNKAYVPEGNSNISKSKATCN---GLFSGTSSTIKKILINVGAGYEDTDAVNAOLKA 345
Db 239 nsrpaytpntqaldpk-fqatntkagplsignsikrkilnvgagvnxktadvnvaqla 297
QY 346 VENLAK-ROITFKDDNGTGKVKLGETLITIKG-ETQADKLTDDNNIGVYTD-NNTGLK 402
Db 298 vkwakerrittfggdnstcdvkgidntltikgaetna--ltd-nnlgvkvkeadnsglk 354
QY 403 VKLAKNLGSETVSTKNLTASEKVVVGSNN-TAELOSGLTFT-PTTNA-STDRTVYGT 459
Db 355 vklaktlnlntevnttlnattvkvgsasstaelldsdslftqntqsgstaktvvgv 414
QY 460 DGLKFTDNS-NTALEDTTRITTKDIGFSNKAGTVDENK-PYLDKDKLKVGNSTLNGGLT 517
Db 415 ngvklfnmaettaaigttritrtdkigfa-rdgvdvdekqapyldkklkvgsv----alt 468
QY 518 VNNTIGGSNKQIOVGADGKIFAD-VNVVNSNAKFFTRITEEIGFADAGVKDKSPY 576
Db 469 idngidagnkkslnlakgssandavtieqlkaak--ptlnagagisvvtetelsvdksgn 526
QY 577 LDKKQLQVGVGKTKDSGINAGDQKISNVKDATDDTDAVTYKQ----LKQVQDADGALQ 632
Db 527 vtaptyni-gvkkteinsdgttsdk--fsvkgsgttnnslvtaehlasylnevnrtadsaiq 583
QY 633 SFSIRDEKQGEFTISNLYSNGNTPN--TFETITFAGENGISINDIAKGVKVGIDP 690
```



```
Db 584 stvkeedddanaaitv-akdttknagavsiilkkgkngltvatk-kdgtvfglsqdsq 641
QY 691 LTPPKLTVGSD---KDGKTLQVIEQVAGNDTKNT- 722
Db 642 ltigkstinndglvtkdnteq-iquvangikftnvngsnpgtgiantariirdkigfag 699
QY 723 -----IRGLSPTLPSTINAGGVRTTEQGN 747
Db 700 sgvavdnkpyldqdklqvgnvkitntgtnaggaikgtlpsltadqss-rnieignt 758
QY 748 IYSDDEKSAASGDTLNTGFLNKLNNNSVGFVSTYNTVDFIDGNATTAKYVYDETNQTS 807
Db 759 i-qdkksnaasindlntcgfnknnnpdlfstydlvdfangnattatvthdntanks 817
QY 808 KVTYDVNVDPEKTIETG-DNGKTKNKGVKTTTLTTNANGK-ATNF--STVNDNALVNAK 863
Db 818 kvvydvrvvddtthigttdnkn--klgvtktnlksangntatnfnvnssdedalvna 875
QY 864 DIAENLTLAKEIHTTKGTADTALOTFKYK---DGATDDEITVYKDGTOQNGKTVNTL 919
Db 876 diaenlntlakeihtekgtadaltqftvkvdennnaddanaityvgqkann--qvntl 933
QY 920 KLKGENGLVATNKDGTVTFGINTQSLKAGDSTTLNKDGLSTIKNPASNEQIQVGADGVK 979
Db 934 tlkgenglnlktkngtvtfginttsglkagks-tlndgglsiknptgseqiqvgadvk 992
QY 980 FAKV-DKGNSSGTCIDGTSRTIKDOIGFTGANGSLDPTKPHLTDKLKVGEVEITNTGNA 1038
Db 993 fakvnnngvvagidgttridtrdeligtngslkskphlskd-----glna 1040
QY 1039 GKKITNIQSGDITQNSDVAVTGGRVYDLKTELESKINSAAQNSHEFSVADPQGNH 1098
Db 1041 gkkitniqsgelagshdavgtkyldiktelenkissstaktagnslhefsvadeqgn 1100
QY 1099 FTVSNPYSYDTSKSDVTFAGENCITTKVKNKGVVRVIGIDQTKGTLTKLTVGNNGKG 1158
Db 1101 ftsvnpysydsksdvtfagengitbkvknvvrvgldqtkgtltpkltvgnnngkg 1160
QY 1159 IVTDSKDGQNTITGLSNTLANVTND--GAGHALSQS--LANDTOKTRAASITGDVLNAGFNILQ 1216
Db 1161 ividsqngqntitglsntlanvntdkgsvrtteqgnlikdedktraasivdvlseagfnlq 1220
QY 1217 GNGEAVDFVSTYDVFIDGNATTAKTVDTSKTSKVYDVNDVNDKTTIEVTSDKKLGVK 1276
Db 1221 gngeavdfvstydvtnfadgnattaktvydtsktskvvydvndvndttiev-kdkklgvk 1279
QY 1277 TTTLTKTSTANGATKE--SAADGDALVKASDIATHLNTLAGDIQTAKGASQASSASYV 1333
Db 1280 ttltstgtgan--kfalsnqatgdalvkasdlvahntlsgdiqtakgasqanssagyv 1337
QY 1334 DADGNKVTYDSTDYKYQVNDKGVDPKNEVAKDKLVAQAQTPDGTLAGMNVKSVINKEQ 1393
Db 1338 dadgnkviydstdnkyyqakngdvtoktevakdklivaqaqtpdgtlagmnavksvinkeq 1397
QY 1394 VNDANKQGINEDNAPIKGLENAKDKTKYNAAVTVGDNLNAVAQTPLTTFAGDTGTTAKKL 1453
Db 1398 vndankkqginednafvkglekaasdnktnaavtvgdlnavaqtplttfagdtgttakk 1457
QY 1454 GETLTIKGGOTDNKLTNDNIGVAVAGTDGTYKLAIDLNLNSVNAGGTRIDEKGSFVD 1513
Db 1458 getltikggotdnkltndnigvavagldgftvklakdlnlhnsvnaggtkldkkgvsfvd 1517
QY 1514 ANGOAKANTPVLISANGLIDLGGKRISNIGAADDNDVAFKQFNEVAKTVNNLNNQNSGA 1573
Db 1518 ssqakantpvlisangldl----- 1536
QY 1574 SLFPVYTDANGKPIINGTDGKPQKAIKAGDKYYHANANGVPVVDKDGKPIPTDADKLANLAA 1633
Db 1537 ----- 1536
QY 1634 HGKPLDAGHVVASLGGNSDAITLTNIKSTLPQIDTPTNTGNANAGQAQSLPSLSAAQSN 1693
```

```
Db 1537 ----- 1536
QY 1694 AASVKDVLNVGFNLQNHNOVDFVKAYDTVNFVNGTGADITSVRSADGTMNTVNTALA 1753
Db 1537 ----- 1536
QY 1754 ATDDGCVLIIKAKDGKFKYRADDLMPNGSLKAGKSADAKTPTGLSLVNPWAGKSGTDAV 1813
Db 1537 ----- 1536
QY 1814 ALNNLSKAVFKSKDGTTTTTSVSDGISIQCKDNSSITLSKDLNGLNVGKVLISNVGKTKDFT 1873
Db 1537 ----- 1551
QY 1874 DAANVQQLNEVRNLGLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVTKAGTVLGGKG 1933
Db 1552 daanvqqlnevrnlglgnagnadnagnvniadiKKdpnsgssnrtvtkagtvlgkg 1611
QY 1934 NNDTEKLATGGVQGVYDKDGNANGDLSNVVWTKQKDGSKKALLATYNAAGQTYLNNPA 1993
Db 1612 nndteklatggiqugvdkdgnangdlsnvwwtkqkgdkgskalllatynaagqtnyltnnpa 1671
QY 1994 EADIRINEQIRFFHYNDGNQEPVVOGRNGIDSSASGKHSVAIGFQAKADGEAAVAIGRQ 2053
Db 1672 eadirinedqirffhvdngnqepvvqgrngidsasgkhsvaigfqakadgeaavaigrq 1731
QY 2054 TOAGNOSIAIGDNAQATGDSQSTAIQTGNVVTGKHSIGAIGDPSTVKADNSYSVGNNOQFID 2113
Db 1732 tqagnosiaigdnaqatgdsqstaiqtgnvvvaghkhsaigdpstvkadnsysvgnnqftd 1791
QY 2114 ATQTDVFGVGNNTVTTSNSVALGNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKG 2173
Db 1792 atqtdvfgvgnntvttsnsvalgnsaisagthagTqakksdgtagtttagatgtvk 1851
QY 2174 FAGQTAVGAVSVCASCAERRIQNVVAGEVSATSTDAVNGSOLYKATQGIANATNELDHR 2233
Db 1852 fagqtavgavsvgasgaerriqnvaagevsatstdavngsqlykatqgsianatneidhr 1911
QY 2234 HQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALVGLSLSDNGQWVFK 2293
Db 1912 hqenkanagissamamampqayipgrsmvtggiathngqgavavglkslndngqwwfk 1971
QY 2294 INGSADTQGHVGAAGAGPHF 2314
Db 1972 ingtsadtqghvgaavgagfhf 1992
RESULT 5
AAW04505
ID AAW04505 standard; Protein; 1992 AA.
XX
AC AAW04505;
XX
DT 25-JAN-1997 (first entry)
XX
DE Moraxella 200 kDa outer membrane protein.
XX
KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;
KW diagnosis.
XX
OS Mycobacterium catarrhalis strain 4223.
XX
PN W09634960-A1.
XX
PD 07-NOV-1996.
XX
PF 29-APR-1996; 96WO-CA00264.
XX
PR 26-MAR-1996; 96US-0621944.
PR 01-MAY-1995; 95US-0431718.
PR 07-JUN-1995; 95US-0478370.
XX
PA (CONN-) CONNAUGHT LAB LTD.
```


Db 1537 ----- 1536
QY 1754 ATDDGNVLIIKAKDKFYKADLMPNGSLKAGKSASDAKPTTGLSLVNPNAKGSTGDV 1813
Db 1537 ----- 1536
QY 1814 ALNNLSKAVFKSDGTTTTTSSDGSISIOGKDNSSITLSKDGVLNVGKVIYSNVGKTKDT 1873
Db 1537 ----- 1551
QY 1874 DAANVOOLNEVRLLGLGNAGNDADNOYNIADIKDPNSGSSSNRTVTKACTVLGGKG 1933
Db 1552 daanvqinevrllglgnagnadnqnvniadiikdpnsgsssnrtv.kagtvlggkg 1611
QY 1934 NNDEKATGCGVGVGDKNANGDLSNVWVKQDKGSKKALLATYNAACOTYLTNNPA 1993
Db 1612 ndteklatggiqgvdkdgnangdlsnvvvtqkdgskallatynaagatnyltmpa 1671
QY 1994 EADRINEQIRFHHVNDGNQEPVVOGRNGIDSSAGKHSAVIGFQAKADGEAAVIGRQ 2053
Db 1672 eaidrineqgirffhvnqepvvggrngidsasgkhsaigfqakadgeaavaiqrq 1731
QY 2054 TORAGNOSIAGDQAQTDGDSIAIGTGNVYTKHSGAIGDPSTVKADNSYSGVNNQFID 2113
Db 1732 tqagndsiaigdnhaqatgdqdsiaigtgnvvagkhsaigdpstvkadnsysvgnnqftd 1791
QY 2114 ATQDVFVGNNITVTFESNVALGSNSAISAGTHAGTQAKKSDGTAGTTTGTAGATGVKG 2173
Db 1792 atqtdvfvgnnitvtfesnvalgsnsaisagthagtqakksdgtagtttagatgtvkg 1851
QY 2174 FAGQTAGVAGSVGASGAERIONVAGEVSATSDVANGSQLYKATQGIANATNEIDHRI 2233
Db 1852 fagqtavagsvgsgaerionvagevsatstdavnsgsqlykatqgs.ianaatneidhri 1911
QY 2234 HONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGA VAVAGLSKLSNDNGQWVFK 2293
Db 1912 hqnenkanagissamamasmpqayipgrsmvtggtiathngqgavavaglsklsndngqvwfk 1971
QY 2294 INGSADTQGHVGAAGAGPHF 2314
Db 1972 ingssadtqghvgaavgagfhf 1992
RESULT 6
ID AAB69137
XX AAB69137 standard; Protein; 1992 AA.
AC AAB69137;
XX 24-APR-2001 (first entry)
XX M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.
XX Moraxella catarrhalis strain Q8; major outer membrane protein;
KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
KW otitis media; detection.
XX Moraxella catarrhalis.
OS WC200107619-A1.
XX 01-FEB-2001.
XX 26-JUL-2000; 2000WO-CA00870.
XX 27-JUL-1999; 99US-0361619.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Sasaki K, Yang Y, Klein MH;
PI WPI; 2001-159722/16.
XX DB

DR N-PSDB; AAF59106.
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis
XX
PS Claim 1; Fig 8A-V; 247pp; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein (II) has antibacterial activity.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis M56 200kDa protein in pKS348, which is given in the
CC exemplification of the present invention.
XX
SQ Sequence 1992 AA;
Query Match 54.8%; Score 6404.5; DB 22; Length 1992;
Best Local Similarity 60.4%; Pred. No. 5.6e-275;
Matches 1427; Conservative 147; Mismatches 316; Indels 471; Gaps 48;
QY 56 VIGATLNGSYAQ--NNSKIAFGTTG---NNDNASANEASIAIGSLAKAHANQAIAGG 110
Db 1 migatlngsyayqkdkthiaigeqnqprsgtakadgdraiaigenanagggaiaigs 60
QY 111 SKDPDRNOAQNQAGSHAKGKESIAIGDVLAEGDSIAIGSDDLYLDRNSTNSKYPNG- 169
Db 61 snktv-ngsvsldkigtatqgesiaigddvkasgdsiaigsdldhldqngnphkpgt 119
QY 170 LLSTLIQNHVTLRQIRDSNGSQ-KYRRTAAEGHASTAVGAMAYAKAHFANAFGRSTRAG 228
Db 120 lindlinghavlkerrskdndvkyrrttasghastavgamsyaqghfsnagftrataks 179
QY 229 NYSLAVGLTAKAEKGYTIAIGSNQAQINYGALAGATRVLDYIGIALGVGSQLNNNN 288
Db 180 ayslavglataeagqstiaigsdatssslgaialagatraqlgsgialgggs-vvtqsdn 238
QY 289 NNNKAYVPEGNSNIKSSKATGN--GLFSGISSTIRKILINVGAGYEDTDVAVNAOLKA 345
Db 239 nsrpaytpntqaidpk-fqatnntkagplsgnsikrklinvgagvntkdavnavqlea 297
QY 346 VENLAK-RQITFRKDDNGTGKVKKLGELTLTKGG-ETQADKLTDDNNNIGVVTD-NNTGLK 402
Db 298 vkwakerritfgddndstadvkigldntltikggaetna--ltd-nnigvvkeadnsglk 354
QY 403 VKLAKNLSGLETYSTKNLTASEKVTYVSGNN-TAELOSGGLTFT-PTNA-STDKTYYGT 459
Db 355 vkiaaktlnnltvnttlnattvtkvsgsssttaellsdslftqptgsgstsktyvgv 414
QY 460 DGLKFTDNS-NTALEDTTRITKDKIGFSNKAGTVDENK-PYLDKDKLVGNSTLNNGLT 517
Db 415 ngvktfnnaettaalgttrtdrkigfa-rdgdvdekqapylkdkkvgsv-----ait 468
QY 518 VNNITIGGSNKOIOVGADGKIFAD-VNVNVSNAAKFGTTRITEEBIGFADADGKVDKKSXPY 576
Db 469 idngidagnkkslnlagssandavtieqlkaak--ptlnagagisvtptseisvdaksgn 526
QY 577 LDKKQLQVGVGKTKDSGINAGDQKISNVKDATDDTDVATYKQ----LKQVQDADGALQ 632
Db 527 vtaptyni-gvktelnsdgtssdk--fsvksgsgnnsilvtaehlasylnevnrtadsalg 583
QY 633 SFSIRDEKGOFFTISNLYSNGNTPN--TFETITPAGENGISISNDIAKGVKVGIDIPNG 690
Db 584 sftvkeeddddanaitv-akdttnagavsilklknngitvatck-kdgtvtfglsqdsq 641

QY	691	LTPKLTGVSD-----KDGKTLQVIEQVAGSNGDTRNI-----	722	QY	1694	AASVKDVLNWFNLOTHNHQVDFVKAYDTVNFVNGTGADITSVRSADGTMSTNVTNTALA	1753
Db	642	lllgkstlnndgltvkdtneq--lqvganglkftvnngsnpgtgiantaritrkdigfag	699	Db	1537	-----	1536
QY	723	-----IRGLSPTLSITNAGGVRTTEOGNT	747	QY	1754	ATDDGDNVLKAKDGKFKYKADDLMPNGSLKAGKSADAKTPTGTLVLNPNACKGSGTGDVAV	1813
Db	700	sdgavdtnkpyldqdklqvgnvkntntginaggkaigtlsptlpsladqss--rnielgnt	758	Db	1537	-----	1536
QY	748	ITSDEDKSKAASIGDILNTGFLKNNSNVGFVSTYNVDFIDGNATTAKTYDYDETQTS	807	QY	1814	ALNLSKAVFKSKDGTTTTTSVSDGISIQGKDNSSITLSKDLNNGVKVISVNGVGTKDT	1873
Db	759	i-qdkksnaasaindintgfnlknnpnidfstydivdfangnattatvthdtankts	817	Db	1537	-----ggkvisnvvgtkdkt	1551
QY	808	KVTYDVNVDEKTIETG--DNGTKNIGVKVTKTLTTTNANGK--ATNF--STTNDNALVNAK	863	QY	1874	DAANVOOLNEVRNLLGLGNAGDNADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLGGK	1933
Db	818	kvydvvnvddtclhltgtdnkn--klgvkttklntsangntatnfnnsdedadlnvak	875	Db	1552	daanvqqlnevrnlllglnagnadnagvniadikkdpnsyssnrtvikagtcv199kg	1611
QY	864	DTAENLTLAKETHHTKGTADTALOTFKVKK-----DGATDDDETITVVGKDTQNGKTVNTL	919	QY	1934	NNDTEKLATGVOGVVDKGNANGDLSNVWVKTKQDKGSKALLATYNAAGOTNLYLTNNPA	1993
Db	876	dlaenlntlakehttkgtadtalqtfvkkyvdennnaddanaicvgqknann--qvntl	933	Db	1612	nndteklatggigvgvdkdgnangdlsnvwwtkqdkgskalliatynaagqtnylcnnpa	1671
QY	920	KLKGENGLVATNKGDTVTFGINTOSGLKAGDSTTLNKDGLSIKNPASPNEQIQVGADGVK	979	QY	1994	EADIRINEQGIREFHVNDRGNOQEPVVGQGRNGIDSSASGKHSVAIGFOAKADGEAAVIGRQ	2053
Db	934	tlkgenlntkdnktvtfintsglkagks--tldngglsiknptgseqiqvgadgvk	992	Db	1672	ealdrineqgrrfthvndgnqepvvggrngldsaagkhsaigfqakadgeaavaigrq	1731
QY	980	FAKV--DKGNSSTGIDGTSRTITDOIGFTGANGSLDTPKPLTKDKLKVGEVEITNTGINA	1038	QY	2054	TOAGNQSTAIQDQAQNGDOSIAITGTGNVVTGKHSAGIDPSTVKADNSYSVGNNNQFID	2113
Db	993	fakvnnngvvvgagldgtrtrideigtgngslskspklskd-----gina	1040	Db	1732	tqagngslagdnagaqgdgslalgtgnvvagkhsaigdpstcvkadnsysvgnnnqftd	1791
QY	1039	GKKITNIQSGDITQNSDAVTCGRVYDLKTELESKINSAAKTAONSLEHFSVADEQGNH	1098	QY	2114	ATQTDVFGVGNNTVFTESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTACATGTVKG	2173
Db	1041	gkkitniqsgelqaqshdvtgklydltkelenkissstaktagnslhefsvadeqgn	1100	Db	1792	atqtdvfgvgnntvftesnsvalgsnsaisagthagtqakksdgtagtttagatgtvg	1851
QY	1099	FTVSNPYSSYDTSKTSVITFACENGITTKVNGVVRVGGIDQTKLTPKLTGNNNGK	1158	QY	2174	FAGQTAVGANSVGSAGERRIQNYAAGEVSATSTDAVNGSOLYKATOGIANATNELDHRI	2233
Db	1101	ftvsnpyssydtksdviifagengittkvngvvrvgldqckltpkltvgnnmgkg	1160	Db	1852	fagqtavgavsvgsaagerrlqnvaagevsatstdavngseqlykatgsianatneldhri	1911
QY	1159	IVIDSKQGNITIGLSNLTANVND--GAGHALSQG--LANDTDKTRAASIDCVLNAQFNILQ	1216	QY	2234	HONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOCQAVAVGLSKLSDNGQWFK	2293
Db	1161	lvidsqngqntitgltlanvntndkgsrvrttegnlikdedktraasivdvlsagfnlq	1220	Db	1912	hqnenkanagissamamasmpqaylpggrsmvtgglatnngggavavglskldngqwwfk	1971
QY	1217	NGEAVDFVSTYDTPVDIDGNATTAKTYDTSKTSKVYDVNVNDKTEVTSDDKLLGVK	1276	QY	2294	INGSADTQGHVGAAGVAGFHF	2314
Db	1221	nggeavdfvstydvntfnadgnattaktydtdtsktskvvydvnvddttlev--kdkklgvk	1279	Db	1972	ingsadtgghvgaavgagfhf	1992
QY	1277	TTTTTKTSANGNATKF---SAAQDALVKASDIATHLNTLAGDIQTAQKASQSSASVY	1333	RESULT	7		
Db	1280	tttltstgtgan--kfalsnqatgdalvksadlvahIntlsgdiqtakgasqansagvy	1337	AAR99393	standard; Protein; 2353 AA.		
QY	1334	DADGNKYIYDSTDKKYQVNDKGOVDKNKEVAKDKLVAQAQOTPDGTLAOMNVKSVINKEQ	1393	AC	AAR99393;		
Db	1338	dadgnkviydstnkyyqakndgtvdktkvakdklvagaqtpdgtlaqnmvksvinkeq	1397	DT	15-JAN-1997 (first entry)		
QY	1394	VNDANKKGINEDNAFTKGLENAADTKTKNAAVTVGDLNAVAQTPLTFAQDTGTTAKKL	1453	XX	Haemophilus adhesion protein HA2.		
Db	1398	vndankkginednafvkglekasdnktknaavtvgdlnavagtpltfagdlgttakk1	1457	XX	Haemophilus adhesion protein; HA2; hsf protein; vaccine.		
QY	1454	GETLTKGGQDTNKLTDNNIGVWAGTDGFTVKLAKDLTLNLSNVNAGGTRIDEGKISFVD	1513	XX	Haemophilus influenzae type b strain C54.		
Db	1458	getltkggqtdtnkltdnnigvvgagtdgftvklakdltnlnsvnaggtkiddkgsfvfd	1517	XX	WO9630519-A1.		
QY	1514	ANGQAKNTPVLSANGLDLGGKRIISNICAADVDDNDVNFQFNEVAKTVNNLNQNSGA	1573	XX	03-OCT-1996.		
Db	1518	ssgqakantpvlisangldl-----	1536	XX	22-MAR-1996; 96WO-US04031.		
QY	1574	SLPFFVVTDANGKPIINGTDCKPQKAIKGADGKYVHANANGVPVDKDGKPIITDADKLANLAA	1633	XX	24-MAR-1995; 95US-0409995.		
Db	1537	-----	1536	XX	{UYSL-} UNIV ST LOUIS.		
QY	1634	HGKPLDAGHOVVASLGGNSDAITLTNIKSTLPQIDPTNTGNANAGQAQSLPSLAAQOSN	1693	XX	{UNIW } UNIV WASHINGTON.		
Db	1537	-----	1536	XX	Barenkamp SJ, St Gene JW;		
				XX	WPI; 1996-455364/45.		

QY 1493 NLNSVNAGTRIDEKGISFVDANGQAKANTPVLISANGIDL---CGKRTSNIGAAVDDNDA 1549
Db 1599 n-----ngvltetdgkdegnglvtakavidanvkagvrvtktgangqndd- 1644
QY 1550 VNFKQFNEVAKTVNNLNQNSGASLPFVVVDANGKPIGTGDKPQKAIKGA-DGKYHYHA 1608
Db 1645 -----fatvasgtn-----vtfadg---ngttaevtkandgsitvkvnyk 1681
QY 1609 NANGVPVDKGRPIITDADKLANLAHAGK---PLDAGHGVVASLGGNSDAITLTFNIKSTLP 1665
Db 1682 vadgikldgd-kivadttvlt--vadgkvtapnngdgkfvdasgldalnklswtat-- 1736
QY 1666 QIDTPWGNANAGQAOQSLPSLSAAQCSNAASVKDVLNVGFNLQTNHNOVDF-----VK 1718
Db 1737 -----agkegtgevd--pansagvevkgd-kvtfkagdnkikqsgkdftyslkkelk 1787
QY 1719 AYDTNVFVN---CTGADITISVRSADGTSNITV---NTALAAATDDGNNVLKAKDKFYK 1772
Db 1788 dltsvefkandggtgsestki-----tkdglitipangagaagantantistvkdg---- 1838
QY 1773 ADDLMPGSLKAKSASDAKTPTGLSLVNPANAGKSGTGDVALNNLSKAVPKS-----K 1826
Db 1839 -----isagnka-----vtnvsvglkkgdghtlangtvadfekehdydnayk 1879
QY 1827 DGTITTTVSSDGISIQKDNSSITLSKDLNVG---GKVIS---NVGKGTKDPTDAANV 1878
Db 1880 dlt-----nidekgadnmpvtadntaatvgdlrglvgwisadkttgepnqeyn-aqv 1930
QY 1879 QQLNEVRNLGLG-NAGDNADGNVNIADIKDP-----NSGSSSNRTVTKAGTV 1928
Db 1931 rnanevkfsgnginvsqgtlntgtrvttfelakgevvksneftvknadgsetnlkvqgdm 1990
QY 1929 LGKGKGNND--TEKATGGVGVVDKGNANGDLSNVVTKQDGSKKALLATYNAAGQTN 1986
Db 1991 yyskediapatskpmtg----ktekykvengkv-----vsangsktevtlltnksgs--- 2037
QY 1987 YLTNNPAEADIRINEOGIRFFHVNDGNQEPVVOGRNGIDSSASGKHSVAIGFOAK----- 2041
Db 2038 yvtgn--qvadaiaaksgfel-----gladaaeakafaesakdkqlskd 2079
QY 2042 -----ADGEAAVAIGRQTQAGNQSTAGDNAQATGDQSTAGTGNVNVTKHSGAIGDPS 2095
Db 2080 kaetvnaahdkvrfanglntkv---saatvestdangok-----vtt-----t 2118
QY 2096 TVKADNSYSVGNNNQFIDATQTDVFGV---GNNITVT-----ESNSVALGSNSAISAG- 2145
Db 2119 fvktd-----velptqlyntdangnkivkkadgkwyelnadgtasnkvtlgn 2167
QY 2146 --THAGTOAKKSDGTAGTITTTAGATGTVKGFAGTAVGAVSVG-----ASGAE 2191
Db 2168 vdangkkvkvvtengadkwyvntnadgaadkktgevsndkystdekhhvrlidpnnqngky 2227
QY 2192 RRIONVAAGESVATSDAVNGSQLYKATQGTIAN---ATNELDHRHONENKANAGITSSAM 2248
Db 2228 vvidhvangeisaatdaingsqlayavakvtnlagvnnlegkvkvkgradagtasal 2287
QY 2249 AMASMPQAYIPGRSMVTGIIATHNGQGAVALGSLKSDNGQWFKINGSADTQGHVCAAV 2308
Db 2288 aasqlpqatmpgksmvalagssyggqglaigvarisdngkviirlsgttnsqgktgva 2347
QY 2309 GAGPHF 2314
Db 2348 gvgyqg 2353

RESULT 8
ID AAB23860 standard; Protein; 2411 AA.
XX
AC AAB23860;
XX
DT 17-JAN-2001 (first entry)

XX Haemophilus influenzae adhesin (Hia) protein from type c strain API.
DE Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
XX non-typeable Haemophilus influenzae; antinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX Haemophilus influenzae.
XX WO200055191-A2.
PN 21-SEP-2000.
PD 16-MAR-2000; 2000WO-CA00289.
PF 16-MAR-1999; 99US-0268347.
XX (CONN-) CONNAUGHT LAB LTD.
PA Loosmore SM, Yang Y, Klein MH;
XX WPT: 2000-618897/59.
XX N-PSDB; AAA92499.
XX Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Haemophilus influenzae
PT infection
XX Claim 1; Fig 24; 275pp; English.
PS The present sequence represents a Haemophilus influenzae adhesin
XX (Hia) protein from the type c Haemophilus influenzae strain API.
CC Hia genes and proteins have antinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX Sequence 2411 AA;
SQ

Query Match 10.1%; Score 1175.5; DB 21; Length 2411;
Best Local Similarity 23.5%; Pred. No. 5.8e-44;
Matches 645; Conservative 369; Mismatches 963; Indels 771; Gaps 127;
QY 1 MNHIYKVFNKATGTFMAVECAKSHSGSSSTAGQVGS-----P 42
Db 1 mkifnvlwvmtqvwvvtlthkrasatvetavlillfatvqanatedeeldp 60
QY 43 VLRITRVAATL-----AILVIGA-----TLNGS 64
Db 61 vrtapvlshfdsketgkevtensnwqlyfhnkgvllkagaitikagdnllkqstnas 120
QY 65 AVAQN-----NSKIAFGTTGNN-DNASASNEASIAIGSLAKAHANQATAIGGS 111
Db 121 sftyslkkdltdltsvateklsfgangkvdtsdangiklaktngnvnhlh---gldst 177
QY 112 KPD-----PRNOAANOK-----AGSHAKGESIAIGDVLAEGDASTA 149
Db 178 lpdavntgtvlsstsfndvkeatratvkdvlneagvnlkagkt--aggnv----- 226
QY 150 IGSDDLVLDNRNSTNSKYPNGLLSTLQNTHTVLRQIRDSNGSKYRRTAAEGHASTAVGAM 209

Db 227 -esvdlvsayn--nvfeitgdkntl-----dvvltakengtktevkftpktsvikekgkl 279
QY 210 AYAKGFANAFGRSTAGNSYSLAVGL-TAKA-----ERGYTI-AIGSNAQAIFYGALA 261
Db 280 ftgkennndnkvsntatndtdegglvtakavldavnkagvrwktttangqgdftatva 339
QY 262 LGADTRVLDLYGTALGVSQILNNNNN--NNKAYVPEGNGSNIKSKS-----AT 309
Db 340 sg--tnvtfesgd--gftasvtdktngngltvkydakvgdglkfdskkivadtaltvt 395
QY 310 GNGLFSGSTIRKRIINVGAGVEDTDAVNVQALKAVENTL---AKROIIFKGDNGGVK 366
Db 396 ggkvaelaiakeddskklvnag-----dlvt-----alignswkakaeadtdtdgalegis 444
QY 367 K----KLGETLTIKGGE-----TOADKLTDDNNNIGVYDNTGLKVKLAKNL 409
Db 445 kdgevkagetvtfkagknlkvqdganftyslqdalgtlsitlgttngngndaktvink 504
QY 410 SGLE-----TVSTKNLTASEKVTVGSGNNTAELOSGL-----TFPTTINASTDKT 455
Db 505 dgltitpagnngtgtntisvt-kdglkagnkaltinvasglrayddanfvdlvnsatdln 563
QY 456 VYGTDLGLKFTDNSN-----TALEDTRITKD---KIGF--SNKAGTVBENKPYLBDK 503
Db 564 rhvedaykgllnlnnekanqplvtdestaatvgdlrklgwwvstkngtkesqykgade 623
QY 504 L----KVGNSLT----NNGGJTVNNNTIGGSNKQIQVGADG---IKPADVNVVNSNAKFGT- 553
Db 624 vlftgagaatvtsksengkhtitvsvaetkadsglekdgtklkikvndqntdnlvtvgnn 683
QY 554 -TRITE-----EEIGFPADAD-GKVDKSPYLDKKQLQVGGVKITKDSGINAGDOKISNVK 606
Db 684 gtavtkgfgtvtgtadargkv-----tvkdatanadkdvkatv 725
QY 607 D-ATDDDAVYTKOLQVOQDADGALQSFSSIRDEKGOEFTISNLSNG--NTPNTEFIT 663
Db 726 dvatainsaafvktenltside-----dnptdngkdalkagdtlt 768
QY 664 F-AGEN-----GISINDIAKG-KVKV-----GIDPINGLELT-TPKLTVGSDKDG 704
Db 769 fkaagknlkvrdgknlftldaknlkvaktavsdltlgttngptggttatpkvntitstadg 828
QY 705 KTQLVIEQVAGSNDTKNI-IRGLSPTL--PSIYNAGGVRTTEQGNNTTSDEDSKAASIG 761
Db 829 ln---faketadaagsknvykglattlttpep-----agaksshvdlvndat-k-snaasie 881
QY 762 DILNTEFLNKNNSGVFSVTYNTVDFIDGNATTAFTYDETNTQTSKVTYDVNVYDEKTIE 821
Db 882 dviragwnigngnnvdyatvtdvntfddstgtttv--tqkadgkadvkigaktv 939
QY 822 LTGDNGKTNKIGVKTTLTTLTTNANGKATNFSTTDND-----ALVNAKDIAENLNTLAK 874
Db 940 ikdhngk-----lftgklidkdnangatvsedgkdgtglvltaktvldavnsksw 989
QY 875 EIHTTKTADTALQTFVKVKDGT-----DDEITVKGDTONGKTYNTLKLKGENGLTV 929
Db 990 rvtgagataet-----gatavnagnaetvt-----sgtsvn-----fknagnatta 1029
QY 930 ATNRKDG---TWTFGINTQSGLKAG-----DSTTLNKGDLISKNPA-----SNEQIQVG 974
Db 1030 tvskdngnlnkvdydvngdglkigdkkivadtltltvtgkvsvpagansvnnkklvn 1089
QY 975 ADGVK-----FAKVDK----- 985
Db 1090 aeglatallnswtakakadyadesgetdqvkvkagdkvtfkagknlkvkqsekdfcysl 1149
QY 986 -----GNSSTGIDGTSR-----ITKDQIGFNGANGSLDTPKPLTKKLKVGCEVIT 1032
Db 1150 qdtlgtitslgttngandrntgtvinkdgtltitlangaaagt-----dasngntisvt 1203
QY 1033 NTGINAGKKITNIQSG-----DITQNSNDA-----VTGGRVYDLKTELESKINSAA 1079
Db 1204 kdgisagnkeitnvksalkkykdcqntagatqpaantaevakqdlvldtkpatgaagnga 1263

QY 1080 KTAQNSLHEFSVADEOQHFTVSNPYSSYDTS-----KTSDVITTFAGENGITTKV-- 1129
Db 1264 dakapdttaatvgdlrglgwvlsaakktadetqkdfhaavknanavefvkgngatvsakt 1323
QY 1130 -NKGVVRVGIDQTK-----GLTTP-----KLTVCNNNGKGV-IDSKDQGNITIGLSN-- 1175
Db 1324 dnnghkvtidvaeakvgdglektdgkiklvdntdgnnltlvtatkgasvakgefnav 1383
QY 1176 ---TLANVTN-DGAGHALSQG-----LANDTKTRAASIGDVLNAGFNLOGNEAVDFVS 1236
Db 1384 ttdattaggttneargkvvvvkgsgngatataetdkkvtatvgdvaka----- 1428
QY 1227 TYDVFDFIDGNATTAFTYDGT-----SKTSKVYVDVNVDMKTIETVSDKKLGVKTTTLT 1281
Db 1429 -----indaatfvkveddsatisiddspiddgandalkagdtltlkgknlkvk----- 1476
QY 1282 KTSANGNATFESADGDALVKASDIATHLNTLAGDIQTAKGASQASASSASVVDADGNKVI 1341
Db 1477 ---rdgknitfalan-dlsvksatvskls-----lgtnngkvn 1511
QY 1342 YDSTDKKYYQVNDKGQVDKNKEVAKDKLVAQAQTPDGTLAQMNKVSINKVEQVNDANKQ 1401
Db 1512 itsdtkglnfakd-----sktgddanihlnhnglasiatldtllnsgattnlngngidnekr 1567
QY 1402 GINEDNAP-----IKGLENAAKDTTKTNAAVTVGDLNVAQOTPLTFAGDTGTAKKLGET 1456
Db 1568 aasvkdvlngagwnvrgvkpasannqveni-----dfvatydtvdfvsgdkdt-----s 1616
QY 1457 LTIKGGOTDYNKLTDDNIG---VVAGTDG--FVVKLAKDLTLNLSNVNAGGTRIDEKGIS 1510
Db 1617 vttv--skdngkrtevkigaktvskhngklftgkelkdann-----ngvt 1661
QY 1511 FVDANGAKANTPVLSSANGLDL---GGKRISNIGAAVDDNDVAFKQFNEVAKTVNNLNN 1567
Db 1662 vtetdkgdegnglvtakavidavnkagvrvtktgangqndd-----fatvasgtn--- 1711
QY 1568 QSNSGASLPFVVYTDANGKPIGTGPKOKAIKGA-DGKYHYHANANGVPVVDKGPITDAD 1626
Db 1712 -----vtfadg---ngttaevtkandgsiltvkynvkvadgikldgd-kivadt 1756
QY 1627 KLANLAHAGK---PLDAGHQVVASLGLGNSDAITLTNLTSTLPQIDTPNTGNANAGQAQSL 1683
Db 1757 vlt--vadgkvtaepngdgkfkvdasgladalnklswat-----agkegtgevd-- 1804
QY 1684 PSLSAAQOSNAASVKDVLNVGNFLOTNHNQVDF-----VKAYDTVNFVN---GTGADI 1733
Db 1805 pansaggevkagd-kvtfkagdnhlikqsgkdfkyslkelkdltsvefkdganggttges 1863
QY 1734 TSVRSADGTSNITV---NTALAAATDDGNYLIRAKDGKFFYKADDLMPNGSLKAGKSASD 1790
Db 1864 tkf-----tkdgltitpangagagasantantitsvtkdg-----isagnka-- 1903
QY 1791 AKPTGLSLVNPNAKGSTGDAVALNLSKAVFKS-----KDGTTTTVSSDSDISIQGK 1844
Db 1904 -----vtnvsvglkfgdghtlangtvadfehydnyakdlit-----nldeka 1947
QY 1845 DNSSITLSKDLNLYG-----GKVIS---NVGKGTDKTDAANVOQLNEVRNLLGLG-NAGN 1895
Db 1948 dnnptvadntaatvgdlrglgwvisadkttgepnqeyn-aqvrnanevkfksgnginvsq 2006
QY 1896 DNADGNQVNTADIKKDP-----NSGSSSNRTVTKAGTVLGGKGNND---TEKLATGG 1944
Db 2007 ktlngtrvltfelakgevvksneftvnadgselnlkvkgdmyyskedidpatskpmgtg- 2065
QY 1945 VOVGVDKDNANGDLNNVWVKTKQDGSKKALLATYNAAGQTNLTNNPAEALDRINQGI 2004
Db 2066 ---ktekykvengkv-----vsangsktevtltnksg---yvtgn--qvadaiaksqf 2111
QY 2005 RFFHVNDGNEPVVQGRNGIDSSAGKSHSVAIGFOAK-----ADGEAAVAIGRQ 2053
Db 2112 el-----gladaaeaka faesakdkqlskdkaetvnahtkdvfrangln 2155

CC influenza species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
S0 Sequence 1104 AA;

Query Match	8.3%; Score 968; DB 21; Length 1104;
Best Local Similarity	24.6%; Pred. NO. 3.3e-35; Indels 590; Gaps 55;

Qy	813	VNDEKTIELTDGNGKTNKIGVKTTLTTLTT	-----NANGKATNESTTON-----	856
Db	17	vvvseltrahkcasatavavatalasataeanntsvtnglnaygd-nfnfttnnsia	75	
Qy	857	DALVNARKDAENLNTLAKIEHTTKGTADTALQTFKVKDGTADDEITVG-	906	
Db	76	dlekhvqdayglinlnek-----dnksasflv-----adnaeatvgnlrklglwls	122	
Qy	907	-KDTQNGKT-----VNTLKLKGENGLTV-ATNKDGTVTFGINTQSLKAGDSTTLNKDG	959	
Db	123	skngtrnkesyqvkqadevlftsgaatavsasskgkhltitsvkgfaevktdatgg	182	
Qy	960	LSIKNPASNEIQVGADGVKFAKVGKNSSTGIDTSRITKDQIGFTGANGSLDITTPHIL	1019	
Db	183	-----qvnadrgkv-----kaedengadvdkvatvdkakaindaatfcvkves	226	
Qy	1020	TKDLKLVGEVETNTGINAGGKKTINIQSGDITQNSDNVATGGRVYDLKT--ELESKINS	1077	
Db	227	ldddldeng-----aagk-----nettdqalkagdtltikagknikakldq	266	
Qy	1078	AAKTAAQNSLHEFSVADQGNHFTVSNPYSYDTSKTSVDITPAGENGITTKVNKGVVVRVG	1137	
Db	267	ngksv-----cfalakd-----ldvtsakvsdkls-----igdktnkvd	300	
Qy	1138	I-DQTKGLTPEKLVGNNGKIGVIDSKQONTITGLSNLTANVTNDGAGHALSGOLAND	1196	
Db	301	itsdangliklak--tgnong-----qngnvhnglastltd-tltgmttgasngvavq	350	
Qy	1197	TDKTRAASIGDVLNAGFNLOQNGBAVDFVSTYDVFIDGNATTAKVTVDDTSKTSKVY	1256	
Db	351	-nhnraasvadvlnagwnlqngasvdfvnaydvdvfnnglcnvnvttdtahkktvrv	409	
Qy	1257	DVNVNDKLTIEVTSDDKLGKVTTLTKTTSANGNATKFSAADGDALVKASDIATHLNTLAGD	1316	
Db	410	dv-----	411	
Qy	1317	IQTAKGASOASSSAYVDADGNKVYYDSTDKKYYQVNDKGQVDKNEKAVAKDLVAQAOTP	1376	
Db	412	-----tgipgvtytedgktvv--kvdnakyvkaqdgasdmkvv-----e	449	
Qy	1377	DGTLAOMNYK--SVINKEQVNDANKKQGINEDNAF-----IKGLENAAKDTKTNAAVTVG	1430	
Db	450	ngelaktkkvllvasagqnpvklsnvaegteendavsfkqkalqekqvtltasanyangg	509	
Qy	1431	DLNAVAQTPLTPTAGDTGTAKKLGETLTIKGGOTPDNKLTDNNIIVGVAGVDGFTVKLAKD	1490	
Db	510	n-----dadggkatqtlnglnlufkfstdgell---nlkvendvtvtfvpk---	551	
Qy	1491	LTNLNSVYAGGTRIDKSGISFVDANGQAKANTPYLSANGLDLGKRSINTGAAVDNDVAV	1550	
Db	552	-----kgsqvgedgkatqngtktdgdl-----	575	
Qy	1551	NFKQFNEVAKTVNNLNQNSGASLPFVVYTDANGKPINGTDBGKPKQAIKAGDGKYYHANA	1610	
Db	576	--veaselveslnklgw-----	591	
Qy	1611	NGVPVDDGKPIITDADKLANLAAHGKPLDAGHVVASLGGNSDAITLTNIXSTLPIQIDTP	1670	
Db	592	--vgvdkdggeld-----	618	
Qy	1671	NTGNANAGQAOSLPSLSAAQOSNAASVKVDYLVNGENLOTNHOVDFVXAYDTVNFVNGTG	1730	
Db	619	-----lragenlkvkq-----dgtntfyalx	639	

```
QY      2054  TQAGNQSIAIGDNAOATGDQSTAICTGNVVTGKHSGAIGDPSTVKADNSYSVNNNQFID 2113
       :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      2156  tkv---saatvestdangdk-----vtl-----tfvkld-----ve 2183

QY      2114  ATOTDVEGV---GNNTIVT-----ENSVALGSNSAISAG---THACTQAOKSDGTAGT 2161
       :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      2184  lptqiyntdangnklvkkadgkwyelnadgtasnkvtlgnydangkkvvkvteagdk 2243

QY      2162  TTTTAGTGTVGVKFGAGTAVGASVG-----ASGAERRIIONVAAGEVSATSTDA 2209
       :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      2244  wyytnadgaadckkgevsndkstdekhhvrlbpnnqngkgvyidnvangeisatstda 2303

QY      2210  VNGSQLYKATQGIAN---ATNELDHRIHQENKANAGISSAMAMASPMOAYIPGRSMVTG 2266
       :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      2304  lngsqlyavakvgvlnlagvnnledkvknkvkradagtasalaasqjlpqaumpgksmvai 2363

QY      2267  GIATHNGOGAVAVGLSKLDSNGOWVFKNINGSADTQGHVGAAGVAGFHIF 2314
       :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
```

RESULTS

XX	AB23856	standard; Protein; 1104 AA.
XX	AAB23856;	
XX	17-JAN-2001	(first entry)
XX	Haemophilus influenzae adhesin (Hia) protein from NTHi strain 29.	
XX	Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;	
XX	non-typeable Haemophilus influenzae; antiinflammatory; auditory;	
XX	antibacterial; meningitis; epiglottitis; septicaemia; otitis media;	
XX	diagnosis; immunogenic; antigen.	
XX	Haemophilus influenzae.	
XX	WO2000055191-A2.	
XX	21-SEP-2000.	
XX	16-MAR-2000; 2000WO-CA00289.	
XX	16-MAR-1999; 99US-0268347.	
XX	(CONN-) CONNAUGHT LAB LTD.	
XX	Loosmore SM, Yang Y, Klein MH;	
XX	WPI: 2000-618897/59.	
XX	N-PSDB; AAA92495.	
XX	Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for	
XX	use as antigens and vaccines and for treating Hemophilus influenzae	
XX	infection	
XX	Claim 1: Fig 20: 275pp; English.	
XX	The present sequence represents a Haemophilus influenzae adhesin (Hia)	
XX	protein from the non-typeable Haemophilus influenzae (NTHi) strain 29.	
XX	Hia genes and proteins have antiinflammatory, auditory and antibacterial	
XX	activities, and can be used in the production of a vaccine. An	
XX	immunogenic composition comprising an Hia gene, a polypeptide encoded	
XX	by an Hia gene, or a recombinant Hia polypeptide is useful for inducing	
XX	protection against disease caused by Haemophilus strains in a	
XX	susceptible host, preferably a human. An Hia protein is useful as an	
XX	antigen in immunogenic preparations including vaccines, as a carrier	
XX	for other immunogens, and in the generation of diagnostic reagents. Hia	
XX	is useful for treating diseases caused by the infection of Haemophilus	
XX	influenzae such as meningitis, epiglottitis, septicaemia and otitis	
XX	media. Recombinant production of Hia favours high recovery of the	
XX	protein compared to the low recovery of native protein from Haemophilus	

Db 412 -----tqlpvqvrtedktvv--kvdnkyyeakqdgsgadmdkkv-----e 449
 QY 1377 DGTLAQMNVK--SVZLNKEOVNDANKKQGINEDNAF---IKGLENAKADTKTKNAAVTVG 1430
 Db 450 ngelaktkvkvlvsasgqnpvkslnvaeagteendavsfkqkalqekqvlttasnayangg 509
 QY 1431 DLNVAOTPLTITAGDTGTTAKKLGELTKIKGGQTDNKLTDNNIGVAVAGDTGFTVKIAKD 1490
 Db 510 n-----dadggkatqlinglnfklfstgdgell---nlkventdtvftpk--- 551
 QY 1491 LTNLSNVNAGTRIDEKGFISFDANGQAKANTPVLUSANGLDGGKRSINIGAAVDDNDVAV 1550
 Db 552 -----kgsqvqgedgkatigngtkttdgl----- 575
 QY 1551 NFQPFNEVAKTVNNLNQNSQASLPFFVVDANGKPIINGTDGKPKQAIKAGDKRYIHANA 1610
 Db 576 --veaselvesinklgw----- 591
 QY 1611 NGVPVDKGPITPDADKLANLAHAKPLDAGHOVVASLGGNSDAITLNIKSTLPQIDTP 1670
 Db 592 --vgvdkdsgeld-----gasnetlvksqdkvt----- 618
 QY 1671 NTGNANAGQASLPISLAAQSNAAASVKDVLNVGENLOTNHNOVDVFKAYDITVNFVNGTG 1730
 Db 619 -----lkagenlkvkq-----dgtntfyalk 639
 QY 1731 ADITSVRSADGTSNITNTALTAAATDDGNVLKAKDGRFYKADDLMPNGSLKAGKSASD 1790
 Db 640 delvgvks-----vefkd-----tangst-----ngast 662
 QY 1791 AKTPTGLSLVNPAGKSGTGDVAVALNLSKAVFKSKDGTITTTTSSDGSIOCKDNSSIT 1850
 Db 663 kitdkgtitsangangaa-----atdaeklk 689
 QY 1851 LSKDGLNVGGKVIENVGKGTKTDAAVNOQLNEVRNLL-----GLGNAGDNADGN 1901
 Db 690 vasdgisagnkavknvvgslkfgdanfnlptasadnltkqydaykgtlnidekqadkq 749
 QY 1902 QVNIADIKKDPNSSSSRRFTIKAGTVLGGKGNNDTEKLATGGVQGVGDKDGNANGDLN 1961
 Db 750 tiltvad-----nt-----aatv-----gdrlg 766
 QY 1962 V-WKTKQDKGSKALLATYNAAGTNTVLTNPNPAEADRINEQGLRFFHVNDGQEPVVOG 2020
 Db 767 lgwv-isadktteinkeyna-----qvrnanefvfksg 799
 QY 2021 RKGIDSS---ASGKHSVAIGQAKADGEAAVAIGRTOAGNOSIAIGDQAATGDQSIAT 2077
 Db 800 -ngihvsgktvngrrreit--felakdena-----lafgygskalrdntvai 842
 QY 2078 GTGNVVTGKHSIGAIDPSTV--KADNSYSVGNNOFIDATQTDVFGVGNV----- 2125
 Db 843 gtgnvvnnaeksfafgdpnyiedkaggsyafgndnr-itskntfvlngvgnakykangdvd 901
 QY 2126 -----ITV-----TESNSVALGNSAITSAGTHAGTOAKKSDGTAGTITTT 2164
 Db 902 tetvtvkdkgkettvtvpkalgatvensvylgnk---statdkgknlksdgtagnttt 958
 QY 2165 AGATGVKGVFAGQATVAGVSVGASGAERRIQNVAAAGEVSATSDAVNGSQLYKATQGTIAN 2224
 Db 959 agtgtvngfagatahagvsvgaasgeerriqnvaageisatdsdaingsqliyavakgvtn 1018
 QY 2225 ATNELDHRHONENKANAGISSMAMASMPQAYIPCRSMVTGGIATHNGQGAVALGSLK 2284
 Db 1019 ----legqvnkvkradagtasalaasqlpqasmpgksmvsivlagssyagqnglaigvsari 1074
 QY 2285 SDNGOWFEKINGSADTQGHVGAAGVAGPHF 2314
 Db 1075 sdngkviirsgtntsgkktgvaagvggq 1104

RESULT 11
 AAB23857

AA023857 standard; Protein; 1004 AA.
 AAB23857;
 17-JAN-2001 (first entry)
 Haemophilus influenzae adhesin (Hia) protein from NTHi strain M407.
 Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
 non-typeable Haemophilus influenzae; antinflammatory; auditory;
 antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
 diagnosis; immunogenic; antigen.
 Haemophilus influenzae.
 WO2000055191-A2.
 21-SEP-2000.
 16-MAR-2000; 2000WO-CA00289.
 16-MAR-1999; 99US-0268347.
 (CONN-) CONNAUGHT LAB LTD.
 Loosmore SM, Yang Y, Klein MH;
 WPI; 2000-618897/59.
 N-PSDB; AAA92496.
 Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
 use as antigens and vaccines and for treating Hemophilus influenzae
 infection
 Claim 1; Fig 21; 275pp; English.
 The present sequence represents a Haemophilus influenzae adhesin (Hia)
 protein from the non-typeable Haemophilus influenzae (NTHi) strain M407.
 Hia genes and proteins have antinflammatory, auditory and antibacterial
 activities, and can be used in the production of a vaccine. An
 immunogenic composition comprising an Hia gene, a polypeptide encoded
 by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 protection against disease caused by Haemophilus strains in a
 susceptible host, preferably a human. An Hia protein is useful as an
 antigen, in immunogenic preparations including vaccines, as a carrier
 for other immunogens, and in the generation of diagnostic reagents. Hia
 is useful for treating diseases caused by the infection of Haemophilus
 influenzae such as meningitis, epiglottitis, septicaemia and otitis
 media. Recombinant production of Hia favours high recovery of the
 protein compared to the low recovery of native protein from Haemophilus
 influenzae species. A truncated protein has a significantly higher
 amount of recovery than a full-length protein.
 Sequence 1004 AA;

Query Match 8.2%; Score 959.5; DB 21; Length 1004;
 Best Local Similarity 23.7%; Pred. No. 7, 1e-35;
 Matches 373; Conservative 168; Mismatches 336; Indels 697; Gaps 57;

QY 819 TIELTGDNKTKNIGVKYKTTTLTTTNNANGKATNFSTTD-----NDALVNAKOAENLN 870
 Db 46 tvqatttggtstnglkayg-stnnpnfnaagsatdlarqfdgaydglmlnnekdan 104
 QY 871 TLAKETHTTKGADTALQTFVKVKDGATDDEITTVG-----KDGTONKGT---- 915
 Db 105 ll-----vtdokaatvgnlkrklgwvlskngtrneksqqvk 140
 QY 916 -VNTLKLKGENGLTVATNKDG---TWTFGINTOSGLKAGDSTTLNKGDSIKNPASNEOI 971
 Db 141 hadevlfeqkdqgtvtsksengkhtvtf-----tlekd-invknatvskl 185
 QY 972 QVGADGVKFAKVDKGNSSGTIDGTSRTTKDQIGFTGANGSLDTPKPLTKDKLVGEVEI 1031

[illegible]

Db	708	treitfelakdena-----iafsgskalrdntvtaigtgnvvnaeksgafg	753
Qy	2093	DPSTV--KADNSYSVGNQNFIDATQTDVFGVGNNT-----NVTESNSV-----	2134
Db	754	dpyiedkaggsyafondnri---taknifvlgpsvnaekrdangnvlteekervgkdgak	810
Qy	2135	-----ALG---SNSAISAGTHAGTQAK-----KSGDTAGTWTGATGATGVKGFAGTAV	2180
Db	811	tkvtvpqalgetvensvyignastatkdkgnlksdgtagnnttagatgtvngfagatah	870
Qy	2181	GAVSVCASGAEERIONVAAGEVYSATSDAVNGSOLYKATQGIANATNELDHRITHONENKA	2240
Db	871	gavsvasgeerirqlnvageisatstdaingsqlayavakgvcn---lagvknkvgkra	926
Qy	2241	NAGISSAMAMSPQAYIPGRSMVTGGIATHNQGQAVAVGLSKLSNDGQWVFVKINGSADT	2300
Db	927	dagtasalaasqlpqasmpgkmsvsiagsyqgsglaigvrsisdngkvlrlsgttns	986
Qy	2301	QGHVGAAGVAGGPHF	2314
Db	987	qgktgvaagvgvqw	1000
RESULT	12		
AAAB23854			
ID	AAAB23854	standard; Protein; 1002 AA.	
AC	AAAB23854;		
XX			
DT	17-JAN-2001	(first entry)	
DE	Haemophilus influenzae adhesin (Hia) protein from NTH1 strain 33.		
KW	Hia; adhesin; Haemophilus influenzae adhesin; NTH1; infection; vaccine		
KW	non-typeable Haemophilus influenzae; antiinflammatory; auditory;		
KW	bacterial; meningitis; epiglottitis; septicemia; otitis media;		
KW	diagnosis; immunogenic; antigen.		
OS	Haemophilus influenzae.		
XX			
PN	WO200055191-A2.		
XX			
PD	21-SEP-2000.		
XX			
PF	16-MAR-2000; 2000WO-CA00289.		
XX			
PR	16-MAR-1999; 99US-0268347.		
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
XX			
PI	Loosmore SM, Yang Y, Klein MH;		
XX			
DR	WPI; 2000-618897/59.		
DR	N-PSDB; AAA92493.		
XX			
PT	Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for		
PT	use as antigens and vaccines and for treating Hemophilus influenzae		
PT	infection		
XX			
PS	Claim 1; Fig 18; 275pp; English.		
XX			
CC	The present sequence represents a Haemophilus influenzae adhesin (Hia)		
CC	protein from the non-typeable Haemophilus influenzae (NTH1) strain 33;		
CC	Hia genes and proteins have antiinflammatory, auditory and antibacteri		
CC	activities, and can be used in the production of a vaccine. An		
CC	immunogenic composition comprising an Hia gene, a polypeptide encoded		
CC	by an Hia gene, or a recombinant Hia polypeptide is useful for inducing		
CC	protection against disease caused by Haemophilus strains in a		
CC	susceptible host, preferably a human. An Hia protein is useful as an		
CC	antigen, in immunogenic preparations including vaccines, as a carrier		
CC	for other immunogens, and in the generation of diagnostic reagents. Hia		
CC	is useful for treating diseases caused by the infection of Haemophilus		
CC	influenzae such as meningitis, epiglottitis, septicemia and otitis		

media. Recombinant production of H1a favours high recovery of the protein compared to the low recovery of native protein from *Haemophilus influenzae* species. A truncated protein has a significantly higher amount of recovery than a full-length protein.

Sequence 1002 AA;

Query Match 7.7%; Score 895.5; DB 21; Length 1002;
Best Local Similarity 24.9%; Pred. No. 4.7e-32;
Matches 345; Conservative 159; Mismatches 352; Indels 527; Gaps 53;

[illegible]

Db	610	fgdanfnplt-----sadnltkqydnaykgltnldkaskgkgtptvadnt---aat	658
Qy	1928	VLGCKGNNDTEKLTATGCVGVGVDPKDNANGDLSNV--WV--KTQKDGSKALLATTYNAAGQ	1984
Db	659	v-----gdrlgIgwIasdkttgsekeysaqvrnane	690
Qy	1985	TNYLTNNPAEAI DRINBOGIRRFHVHNDGNQEPVQVRNGRIDSSASGKHUISVAGFOAKADG	2044
Db	691	vkfsgnginvsgktldngtr-----eltfelakde	721
Qy	2045	EAAVAIGRQRTQAGNQSTAIGDNAQAATGDSIIAIGTGNVVTVGKHSAGIDPSTV--KADNS	2102
Db	722	na-----lafgsgkalrtdntvaltqgnvvnaeksgafgdpnyiedkaggs	767
Qy	2103	YSVGNNNQFIDATOTDVPFGVGN-----ITV-----TES	2131
Db	768	yafgndlr-ltskntfvlngvgnvnaikyandgdvdtetvtvkdkdgtktvtvpykalgctve	826
Qy	2132	NSVALGSNSAISAGTHAGTQAKKSDGTAGTTTTAGATGTGVKGFAGQATVAGVSVGASGAE	2191
Db	827	nsvylgnk-----statdkgknlksdgtagntttagttgtvngafagatahagvsvgsgce	883
Qy	2192	RRQNTVAAGEVSATSTDVANGVSOLYKATOGIANATNNELDHRIHQHENKANAGISAMAMA	2251
Db	884	rriqnvaagealsatstdalngsqlyavakgvtn----lagqvnykvgrkadagtasalaas	939
Qy	2252	SMPQAYIPGRSMVTGGJATHNGOGAVAVGLSKLSDNGQWVFKNISADTOGHVGAANCAG	2311
Db	940	qlpqasmgsksmvslagsgyqggsgIalgvrisdngkvIlrlsgcttnsqgktgvaaqvgv	999
Qy	2312	FHF 2314	
Db	1000	vqw 1002	

RESULT	13
AAW56322	
ID	AAW56322 standard; Protein; 2039 AA.
XX	
AC	AAW56322;
XX	
DT	19-AUG-1998 (first entry)
XX	
DE	Haemophilus paragallinarum antigenic protein #2.
XX	
KW	Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW	vaccine; chicken infectious coryza; CIC; fowl.
XX	
OS	Haemophilus paragallinarum.
XX	
Key	Location/Qualifiers
PH	
FT	Peptide
FT	I..70
FT	/label= signal
FT	71..2039
FT	/note= "antigenic protein"
XX	
PN	WO9812331-A1.
XX	
PD	26-MAR-1998.
XX	
PF	12-SEP-1997; 97WO-JP03222.
PR	
PR	19-SEP-1996; 96JP-0271408.
PA	(KAGA) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
PA	(KAGA) CHEMO-SERO-THERAPEUTIC RES INST.
XX	
PI	Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX	
DR	WPI; 1998-230318/20.
DR	N-PSDB; AAV22837.
XX	

Thu Sep 13 14:18:07 2001

us-09-361-619-11.rag

```
Db 1585 dgidagnkksinvadgd-----isptsgdvvtgr-----qlyalmk--- 1621
Qy 1696 SVKQVLNVGNFLQNHQVDFVKAYDTVNFVNGNGADITSVRSADGTMSTNVTALAAAT 1755
Db 1622 -----giryvdevsptkqtaptastcggattantagg----- 1657
Qy 1756 DDDGNVLIAKADGKFKYKADDLMPNGSLKAGKSASDAKTPGTGLSLVNPNAKGSTGDAVAL 1815
Db 1658 -----vapgnavatgdia-----ptqpal--pemktalvgdhlav 1690
Qy 1816 NNLXAVPKSKDGTITTTVSSD---GISIQKDNSSITLSKDKGLNKGKVISNVGKCTKD 1872
Db 1691 -plgslkihgdmvkvttisagnqvgislg--pnlsl---ennlvignkpekaklaaqe 1744
Qy 1873 TDAANVQOLNEVRLLGLGNAGN--DNADGNQVNIADIKKDPNS---GSSSNRTVVIKAGTV 1928
Db 1745 gnalvitnkdd-----gnaamvfnncknmvlivlsdkkakpravldgnggaltiv----- 1792
Qy 1929 LGGKGNNDTEKATGGVQVVDKDNANGDLSNVWVTKQDGSKKALLATVYNAAGOTNYL 1988
Db 1793 ----gnddsq--vtlsskkgkldgn--dlslsvtttert-----nadgglekv 1833
Qy 1989 TNNPAAIDIRINEQGLRPFHVNDGNQEPVVOGRNGIDSSAGSKHSVAIGFOAKADGEAAV 2048
Db 1834 etsfaumdd-----glkf-----kadgdkvi 1854
Qy 2049 AIGROTQAGNQSTAGDNGAQAATGDSIAIGTGNVVTGKHSGAIGDPSTVKADNSYSVGN 2108
Db 1855 -----nkninetvel-----vgdenvt-----siddnkvkvslnkklaid 1891
Qy 2109 NQFIDATQDVFVGNNITVTESNVALGSNSAISAGTHAGTQAKSKDGTAGTTTTAGAT 2168
Db 1892 evk1pntdpda-qkgdsivi-----nnggihagnkvitgkasd----- 1929
Qy 2169 GTVKGFAGQTAVGAVSVGASGAEPRIONVAAGEVSATSTDAVNGSOLYKATOGCIANATNE 2228
Db 1930 -----dptsavnrgqintvidnvgnnfnq 1953
Qy 2229 LDHRIHONENKANAGISSAMAMASPMQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSNDG 2288
Db 1954 vndrigdltesraglagamatasignvalpgkttisvgtatfkgenavagmsrlsding 2013
Qy 2289 QWVEFKINGSADTQGHVGAAYVAGGFHF 2314
Db 2014 kvgirlsgmstngdkgaamsvgftf 2039

RESULT 14
AAW56319 standard; Protein; 2042 AA.
XX AC AAW56319;
XX DT 19-AUG-1998 (first entry)
XX DE Haemophilus paragallinarum antigenic protein #1.
XX KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
XX KW vaccine; chicken infectious coryza; CIC; fowl.
XX OS Haemophilus paragallinarum.
XX PH Key Location/Qualifiers
XX FT Peptide 1..70
XX FT /label= signal
XX FT Protein 71..2042
XX FT /note= "antigenic protein"
XX PN W09812331-A1.
XX PD 26-MAR-1998.
XX PF 12-SEP-1997; 97WO-Jp03222.
```

```
XX 19-SEP-1996; 96JP-0271408.
XX (KAGA ) 2H KAGAKU & KESSEI RYOHO KENKYUSHO.
PA (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.
XX Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX WPI; 1998-230318/20.
DR N-PSDB; AAV22834.
XX Antigenic polypeptide from Haemophilus paragallinarum induces HI
PT antibody production - and is useful for diagnosis of and preparation
PT of vaccines for chicken infectious coryza
XX Claim 1; Page 51-69; 108pp; Japanese.
XX The present sequence represents an antigenic protein derived from
CC Haemophilus paragallinarum strain A-221. The antigenic protein
CC stimulates the production of HI antibodies in fowl. The protein
CC and DNA coding for it can be used in the preparation of vaccines
CC for the prevention of chicken infectious coryza (CIC). The protein
CC and its antibodies can be used in the diagnosis and treatment of CIC.
XX Sequence 2042 AA:
SQ
Query Match 6.8%; Score 796; DB 19; Length 2042;
Best Local Similarity 20.9%; Pred. NO. 2.7e-27;
Matches 556; Conservative 322; Mismatches 816; Indels 968; Gaps 118;
Qy 1 MNHYIKVIFNKATGTFMAVAEAKS-HSGSSSTAGVGSSPVI-RLTRVATLAILVIG 58
Db 1 mnkvfkikysvkvqemivvselannkdktsqkntbmtaffqlftkctylallnialg 60
Qy 59 ATL-----NGSAYAQNNS-----KTAFGTGNNDN 83
Db 61 aslfpqlanakwlevssvsklsvsqsnvlnpnsqsvgtspqgvaigvgatndr 120
Qy 84 ASASNEASIAIGSLAKAHANQAIAG-CSRKP----- 113
Db 121 -satgalaigvgvgnknetlakdsialgvgaknestapsvltlgkqlarfkelsivmglnay 179
Qy 114 --DPRNOAANKAGASHAKGK-----ESTAIIGDVLAEGDASIAI 150
Db 180 tqldprgtsketrqgsvvigenaksagqsvslgnsksktnsisigagtfaegkssai 239
Qy 151 GSDDLYLDNRSTNSKYP-----NGLLSTLIQN-----HTVLRQIRDSNGSQ----- 191
Db 240 gtdk--isgtkyndklpatawng-tgtvpknsiwdifselymgkqtngrdydtttrdpnk 296
Qy 192 -----KY-----RRTAAEGHASTAVGAMAYAKGHFANA 219
Db 297 peafykfsdfkgkyvntptasptayagkigalagrtlaagemstavgslafaladrsta 356
Qy 220 FGTRS-----TAEGNY-----SLAVGLTAKAKGYTIAIGSNAQAINYGALALG 263
Db 357 mglrsfvakdvvgtaigeestrfakdsvaigntkeasnagmaygykakavggagalaig 416
Qy 264 -----ADPRVDLYGIALGSGQILNNNNNN----- 289
Db 417 tevaagakfnshqtnllqdnnaayatlnnadsddtkgnaltvtqsfdomltnpliv 476
Qy 290 -NNKAYVPECGNGSNIK-----SSKATGNGLFSGSTTKRKLIINVG-AGYEDTDAVNAQ 342
Db 477 senetylttsagaikkattatdsaggknalagstfaskansvalayaadagnafa 536
Qy 343 LKA---VENLAKRQITFKGDDNGCTGVKKLGETLTITKGGETQADKLTNNNNIGVVTNNNT 399
Db 537 lgsysfvessatntiti-----gvgsyak-gknsfl--ggtwastlsdr-----vvlgnst 585
Qy 400 GLKVKLAKNLGLSETVSTKNLTASEKVTYVSGSGNNTAELQSGGLTFTPTTNA-STDKTYVG 458
```


Db 586 sis-----sg-----sqnalai-gvnnvfigndsasslalmgmstiakeakspdsiaig 632
QY 459 TDG---LKFTDNGN---TALEDTTRITKDKIGFSNKAAGTVDENKPYLDKDKLVG-NST 510
Db 633 kearidakdtngltlyqpyvdettr-----afn-----fnessdym-raqamalgnak 681
QY 511 LNNGLTVNNTIGSGNKQIOVGADGTFADVNVNVSNAKFGT-TRITEEIGFADADGK 569
Db 682 vsrg---vgkmetgin-smaiga---yaqatignstalgvsktdytweql-----725
QY 570 VDKKSPVLDKKQLQV-----GGVKLTGDSINAGDQKISNVKDATDQDDAVTYKQLQV 623
Db 726 --eudpwvsegaisptsrgtkvisv---gskgserfirvnlasgsdtdavnaqlktv 779
QY 624 QODADGALQSFSIRDEKQGFETISNLSNGNTPN--TFETITFAGENGISISNDIARGK- 680
Db 780 ee-----rflseinnllngggvkvylsvektningqsg-rvasqirngen 822
QY 681 ----VKVGIDPINGLETPKLTVGSKDKGTQOLVIEQVASGNDYKNIIRIGLSPILPSTNA 736
Db 823 yeryvkl-----ktqilly-----ldargklingekfdqnsi---852
QY 737 GGVRTTEQGNITSDDEKSKAASIGDILLNTGFNLKNSNSVGFVSTYNTYDFIDGNATTA 796
Db 853 -----nkiravvgeleaeysge-----lktasainqvataq-----leqevtln 891
QY 797 KVTYDETNOTSKVTYDVNVDEKTIETLTDGNGKTNKIGVKTTLTTLTNNANGKATNFSTDN 856
Db 892 --nfdkfnq-----yktqienasnad-sarnvggltp 920
QY 857 DALVNAKDIAENLNTLAKEIHTTKGTADTALQTFKVKKGADGDTTITVKGDTQNGKTV 916
Db 921 qiaaqlkannnylndgak-----gqdsiafgwqaktsgangng---lagkqalaignfa 970
QY 917 NTLKLGKENGLTAVTKDGNVTFTGINTQSGLKAGDSFTLANKDGLSINKPNASNQIQVGAD 976
Db 971 ns-----saenaislgtndstsmt-----gavaigkgaivtagg-----kpsiallgqd 1013
QY 977 GVKFAKVDKNGSSTIGDTSRITKDKQIGFTGANGSLDITTKPHLTOKLKVGEVEITWTGI 1036
Db 1014 stvansaistsspmg-----lifnnfagspet-----lgvlsigtagr 1054
QY 1037 NAGGKKTITNQSODITQNSNDNAVTVGGRVY-----DLKTELESKN-----1076
Db 1055 e---rklivnvaagdvsgasteaingsqlyatnfmllskvadsvksnfgnvnlgtdgtitf 1111
QY 1077 -SAAKTAONSLHE-----FSVAD-----EOGNHFTVS--NPYSSYDTS 1111
Db 1112 tnig99qatihdainnvltkglylkadqndptnggqkvvelgnalcisatngwanngvn 1171
QY 1112 -KTSDVITFAGENGITTKVKNKVVRVIGIDQTKGITTTPKLTGVNNGNGKGIVIDSKDGQNTI 1170
Db 1172 yktnnltynsqn-----gtlilgmredp--svkqitaglynttg-----dannk 1214
QY 1171 TGLSNTLIANTNGAGHALSQGLANDTDKTPAASIGDVLNAGNLQNGEAVDFVSTYDT 1230
Db 1215 nqlnnltiqttleatg-----itsvsgstnyagfslga-----ds 1249
QY 1231 VDFIDGNATTAFTYDDTSKTSKVYVDVNVNDKNTIEVTSDDKILGVKTTTLTKTSANGNAT 1290
Db 1250 vtfsg99agcvklis-----gvsdat-----1269
QY 1291 KFSAADGDLALVKASDIATHLNTLAGDIO-TAKGASQASSASVYDADGNKVIVDSTDPKY 1349
Db 1270 ----adtda-atikqvkeyrttlvgnditaaadrsqstngitynlslnkgtvsateek- 1323
QY 1350 YQVNDKQVDKNKEVAKDKLVAAQTPDGTQLAQNVKSVINKSEQVNDANKKQGINEDNAF 1409
Db 1324 -----vvsqgktvyee-----irnaif-----gnif 1343
QY 1410 IKGLENAAKDTKTKNAAVTVGDDLNAAVAQTPLTFAGDNGTTA-----KKLGETLTIKG 1461
Db 1344 tiglidd-----ttlnkinnpadqdlslslsesgknaitglvdvkktnspitve- 1391

QY 1462 GQDTNKLTDNNIGVWAGTDCFTVKLAKD---LTNLSNVNAGGTRIDEKGISFYDANGQA 1518
Db 1392 pstdsnnkkkftvg-vdftdteitgedatddkklttsksve-----syv-tukla 1438
QY 1519 KANTPVLNGLDGGKRIENIGAAVDNDNDAVNFQKQNEVAKTVNNLN-----NQSNSGA 1573
Db 1439 nftsdillisdg-----rsgnattandvgvgrlisd-gftiksenftlgsqyngsd 1488
QY 1574 SLFPVVT'DANGKPI'NGT'DGK'PQKAIKAGDGKYYHANANGVPVDKGPITDADKANLAA 1633
Db 1489 slgvmvddqng-----1499
QY 1634 HGKPLDAGHGVVASLGSNSDAITLTNIKSTLPQIDTPTNTGNANAGQASLPSLSAAQOSN 1693
Db 1500 -----vfklslnmtalt-tslantfakidasn-----ltdsdn 1531
QY 1694 AASVKDLVNGFNLOTHNQVDFKAYDVTNFVNGTGADITSRSADGTMSNITVNTALA 1753
Db 1532 kekwrtnlv-----ysktevdaeiqskvltltpdsglifatkqagsgnnagldagnkkl 1586
QY 1754 ATDDDGNNVLKAKD---GKFYKADDLMPNGSLKAGKASDAKT--PTGLSLVNP- 1803
Db 1587 snvadgdisptsdgvvtgrqlya--lmqgriyvvgdevspktqtaptanptattapt 1644
QY 1804 -----AGKGSTGD-----AVALNLSKAVFK 1824
Db 1645 asstqgwattantaggvapagnvatgdiaptqptlpemmtalvddhlaavlpggskli---1701
QY 1825 SKDGTITTTTSSD---GISIQKDNSSITLSKQGLNVGGKVISNVGKTGDTDANVQOL 1881
Db 1702 hgdhmvkttisadnqvglisq--pnisie-----nnlvigsdnpekakiaa- 1745
QY 1882 NEVRNLGLGNAGNDNA-----DGNOVNIADIKKDPN---SGSSSNRTVIAKACTVLGGK 1932
Db 1746 -qegnalvitrnkdgnaamvfnnekmlvlvsdkeakprvlldgngaltiv-----1795
QY 1933 GNNDTEKLATGGVQGVVDKDNANGDLSNVMVTKQDKGSKALLATYNAAGQTYLTNNP 1992
Db 1796 gnddsq--vtlsskkkgkdidgn---dlslsvtttert-----nadgglekvetsf 1840
QY 1993 AEADRINEGIRFFHVNDNQEPVVOGRNIDSSASGKHSVAIGFOAKADGEAAVAIGR 2052
Db 1841 atmdad-----glkf-----kadgdkvi-----1857
QY 2053 QTOAGNOSIAIGDNAQATGQDSIAIGTGNVVVTGKHSIGAIGDPSTVKADNSYSVGNNOFI 2112
Db 1858 -nkklnetvel-----vgdenvtl-----sitddnkvkvslnkklaiddevki 1898
QY 2113 DATQTDVFGVGNNTITVTSNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTATAGTGVK 2172
Db 1899 pntdpda-qkgsdsvi-----nngghagknkvitgvkasd-----1932
QY 2173 GFAGQTAVAVSVGASGAERRIQNVAAGEVSATSTDAVNSQLYKATQOGIANATNEIDHR 2232
Db 1933 -----dptsavnrqglnvdivdnvqnnfnqnvqr 1960
QY 2233 IHQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAVGSLKSLONGQWVF 2292
Db 1961 igdltresagigamataslqnvalpgkttisvgtatfkgenavaigmsrlsdngkvgl 2020
QY 2293 KINGSADTQGHVGAAGVAGGFHF 2314
Db 2021 rlsqgmstngdkgaamsvgfsf 2042

RESULT 15

AAB23858
ID AAB23858 standard; Protein; 1094 AA.
XX
AC AAB23858;
XX
DT 17-JAN-2001 (first entry)

Db 1 MNHIYKVFENKATGTFMAVEAYAKSHSHGGSCATGQVGVSRVTLSFARIALAVLIVIGAT 60
 Qy 61 LNSAYAQNNSKIPAGTGTGNNDNASASNEASTAIGSLAKAHANQAIAGGSKPDPNRQAA 120
 Db 61 LNSAYAGIGISEADGGK---GANARCDKSAIGADIAQALGSGSIAIGDNK-IVHNSNN 116
 Qy 121 NOKAGSHAKGRESIAIGDVLAEAGDASTAIGSDLLYLDNRNSTNSKYPNGLLSTLIQNHIV 180
 Db 117 NANIKAKASGNEAIAIGDVLASGHASTAIGSDLLYLLKKTVOO---ISELLPIIROOKA 173
 Qy 181 LR---QIRDSNCSOKYRTAAEGHASTAVGAMAYAKGHAFANAPGTRSTAEAGNYSLAYGLT 237
 Db 174 LNDIYOLADTN-LQYRTHAQGHASTAVGAMSAKGFHAFNAGTRATAEGTYSILAVGLT 232
 Qy 238 AKAEGYTTAIGSNAQAINYGALGADTRVDLYGIALGYGSQL---NNNNNNNNKAY 294
 Db 233 AYAKAASSIAGVCSNAQAIAGFAATAVGGSTQVNLNRGIALGFGSOVLQKRDNDVNAANRAY 292
 Qy 295 VPEGNGSNKSKAT-GNG---LFGTSS---TIKRIINVGAGYEYTDVAVNVAQLKAV 346
 Db 293 APDQNPIDNRKATFKNGATDFVSTGNSGNDSTIRKLIINVGASADTDVAVNVAQLKEA 352
 Qy 347 ENLAKQIITFKGDNGTGVKKLIGETLTITKGGETQADKLTDNNNIGVYTDNNTGLKVKLA 406
 Db 353 VRLANQIITFKGDDSNRVEKGLKTLTITGG-AOTSALTD-HNIGVV-QNGDGLKVOLA 409
 Qy 407 KNLGSETVSTKMLTASEKVTVGSGNNTAELQSGGLTFTPTTNASTDXTVYGTDLKFTD 466
 Db 410 ETLTSLKMWNTENLTANEKTVCK----- 433
 Qy 467 NSNTALEDTRITKDKIGFSNAGTVDENKPYLDKD----- 502
 Db 434 -----TRLTDLKIGTNDMNGIDESKPYLDKDTGTHAGGQKITKLTAGVDDDAAT 484
 Qy 503 -----KLGK----- 505
 Db 485 YGOLKKNQTAESALOTFTVKKYDKNGNDANDSKIIITVGNKNNKPDGTQVNTLKLKENGCV 544
 Qy 506 -----VGNSTLNGGLTVNNTIGGSNNKQIOVGADGKIFADVNVN 544
 Db 545 DVTTETNGTFTVGLNQNNGLTVGNSTLNDGSLVKNT--NSNKQIOVGADGKIFFTDISNS 602
 Qy 545 VSNAAKFGTTRITEEIGFADAGKVDKSPYLDKQLQGVGVKITYKDSGINAGDOKISN 604
 Db 603 KPGAGIENTRITRDGIGFANFTGSLDANKPRL-----TPGICINAGGKELTN 649
 Qy 605 VKDATDODTAVTYKQLKQVQDADGALQSPSIR-----DEKGOEFTISNLXNGMTPNT 658
 Db 650 VQSA-----INPATNGGQLDFMNLRLSTANTEKSGSAATIKDLYNLSQVP-- 693
 Qy 659 FETITFAGENGISI-----SNDIAKGVKVGIDPT-NGLTTPKLIVGSD 701
 Db 694 ---LTFAGDTGPNVTKKLGEILKVGKGTADDLTKNIGVAVDSDNSLTVLAKTLSLD 750
 Qy 702 KDG---KTQLVIEQVA--SGNDTKNI----- 722
 Db 751 LDVANTKLTASDKVTVDSGNNATKLONGDLTFSKQNTGATPATNSKTIYGVVDGLKFTDN 810
 Qy 723 -----IRGLS 727
 Db 811 NGIALDGTITYITKDKVGFAGKQDGLSKSPYLDKDKLVGVEITTINGINAGGKAITGLS 870
 Qy 728 PPLPSITNAGGVRTTEQONTITSDEKSKAASIGDILNTGFKNLKNSNSGFFVSTYNTVD 787
 Db 871 NPLTDATNATTHVTVQLG--IVDSTDKTRAASIGDVLNAGFNKLNKNGDAKDFVSTYDITVD 928
 Qy 788 FTIDGNATTAKVYDETNQTSKYTVDVNVDEKTIETLTDGNGKINKIGVKTTLTTLTTNANG- 846
 Db 929 FINGNATTAKVYTD--GRASKVAYDVNVDTGTHLTGADGNKNQIGVKTTLTTLTKTDAKD 986
 Qy 847 KATNFSTT--DNDALVNKADTAENLNTLAKELHTTKGTADTALOTFKVKK-----DCATD 899
 Db 987 KATNFVNSGDDKALINAKDIADNLTLAGEIRNTKGTADTALOTFKVKKENGDDND 1046

Qy 900 DETITVKGDTONGKTVNTLKLKGNGLTVAATNKDGTVTFGINTQSGLKGADSTLLNKDG 959
 Db 1047 ADTITVKGDAKTN--QVNTLKLKGNGLDIQTNKDGTVTFGINTQSGLKGAGNNTLLNNG 1104
 Qy 960 LSTKNPASNEQIOVGADGVKFAKVDKGNSTGIDGTSRITKDDOIGFTGANGSLDTTKPHL 1019
 Db 1105 LSTKNFAGNEQIOVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFAGTNGSLDKSKPHL 1164
 Qy 1020 TKDKLVGVEITNTGINAGGKITINQSGDITQNSNDVATGRRVYDILKTELESINSAA 1079
 Db 1165 SKD-----GINAGGKIIINQSGETAQNSNDVATGKGIYDLKTELENKISSTA 1212
 Qy 1080 KTAQNSLHEFSVADEQGNHFTVSNPYSSYDTSKTSVITFAGENGITTKVNGVVRVGD 1139
 Db 1213 KTAQNSLHEFSVADEQGNFTVSNPYSSYDTSKTSVITFAGENGITTKVNGVVRVGD 1272
 Qy 1140 QTKGLTTPKLTIVGNNGKGVITDSKQONTITGLSNTLANVTND-GAGHALSQG-LANDT 1197
 Db 1273 QTKGLTTPKLTIVGNNGKGVITDSQNGONTITGLSNTLANVTNDKGSVRTTEQGIIRDE 1332
 Qy 1198 DKTRAASIGDVLNAGFNLOQNGEAVDFVSTYDTPVDFIDGNATTAKVYDTSKTSKVYD 1257
 Db 1333 DKTRAASIVDVLNAGFNLOQNGEAVDFVSTYDTPVNEADGNATTAKVYDTSKTSKVYD 1392
 Qy 1258 VVNDKTIETVTSKGLGVKTTTLTKTSANGNATKF---SAADGDALVKASDIATHLNTLA 1314
 Db 1393 VVNDTITIEV-KDKKLGKVTTLTSTGTGAN--KFALSNOQTGDALVKASDIVAHLNTLS 1449
 Qy 1315 GDIOAKGASQASSSASVYDADGNKVIYDSTDKKYQVNDKQVNDKNEKAKDKLVAQAQ 1374
 Db 1450 GDIOAKGASQANSSAGYVDADGNKVIYDSTDKKYQAKNDGTVDKTEKAKDKLVAQAQ 1509
 Qy 1375 TPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFKGLENAAKDTKKNAAVTVGDLNA 1434
 Db 1510 TPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFKGLENAAKDTKKNAAVTVGDLNA 1569
 Qy 1435 VAOTPLTFAGDTGTAKKLGSETLTKGGQDITKGLTNNIGVAVAGTDGFTVKLAKDLTNL 1494
 Db 1570 VAOTPLTFAGDTGTAKKLGSETLTKGGQDITKGLTNNIGVAVAGTDGFTVKLAKDLTNL 1629
 Qy 1495 NSVAGGTRIDKEGISEFVDANGQAKANTPYLSANGLDLGGKRRISNIGAAVDDNDVNFQ 1554
 Db 1630 NSVAGGTRIDKEGISEFVDANGQAKANTPYLSANGLDL----- 1667
 Qy 1555 PNEVAKTVNNLNQNSGASLPFVVTDANGKPIGTGDKPQKAIKGADGKYYHANANGVP 1614
 Db 1668 ----- 1667
 Qy 1615 VDKGKPIITDADKLANLAHKGPLDAGHQVVASLGNSDAITLTNIKSTLPQIDTPNTGN 1674
 Db 1668 ----- 1667
 Qy 1675 ANAGQAQSLPSLSAAQSSNAASVKDVLNNGVFNLOTHNQVDFVKAYDTVNFVNGTGADIT 1734
 Db 1668 ----- 1667
 Qy 1735 SVRSADGTSMTNITVNTALAATDDDDGNVLTAKADGKFFYKADDLMPNGSLKAKSADAKTP 1794
 Db 1668 ----- 1667
 Qy 1795 TGLSLVNPNGAGKSTGDAVALNLSKAVPKSKDGTITTTVSSDGLSIQKDNSSITLSKD 1854
 Db 1668 ----- 1667
 Qy 1855 GLNVGKGVISNVKGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNTIADIKKDPNS 1914
 Db 1668 ----GKVLISNVKGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNTIADIKKDPNS 1723
 Qy 1915 GSSSNRTVITKAGTVLGGKKNNDTEKLTATCGVQGVQVNDKGNANGCDLSNVVWTKQKDGSKKA 1974
 Db 1724 GSSSNRTVITKAGTVLGGKKNNDTEKLTATCGVQGVQVNDKGNANGCDLSNVVWTKQKDGSKKA 1783

QY 1975 LLATYNAAGTNTLTNNPAAIDRIINEQGIREFPHVNDGNOEPVVGNGIDSSASGKHSV 2034
Db 1784 LLATYNAAGTNTLTNNPAAIDRIINEQGIREFPHVNDGNOEPVVGNGIDSSASGKHSV 1843
QY 2035 AIGFOAKADCEAAVATGROTAQGNOSTAIGDQAATGDQSIATGTVNVTGKHSAGLDP 2094
Db 1844 AIGFOAKADCEAAVATGROTAQGNOSTAIGDQAATGDQSIATGTVNVTGKHSAGLDP 1903
QY 2095 STVKADNSYSGVGNNOFIDATQTDVFCVGNNTITVTSNVALGNSAISAGTHAGTQAKK 2154
Db 1904 STVKADNSYSGVGNNOFIDATQTDVFCVGNNTITVTSNVALGNSAISAGTHAGTQAKK 1963
QY 2155 SDGTAGTTTAGATGTVKVPAGTAVGAVSVGASGAERRIQNVAAGEVSATSTDAVNGSQ 2214
Db 1964 SDGTAGTTTAGATGTVKVPAGTAVGAVSVGASGAERRIQNVAAGEVSATSTDAVNGSQ 2023
QY 2215 LYKATOGIANATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQ 2274
Db 2024 LYKATOGIANATNELDHRHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQ 2083
QY 2275 GAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGAGPHF 2314
Db 2084 GAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGAGPHF 2123

RESULT 2

US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33
; TYPE: PRT
; LENGTH: 2353
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 10.1%; Score 1177.5; DB 4; Length 2353;
Best Local Similarity 23.3%; Pred. No. 4.6e-57;
Matches 644; Conservative 349; Mismatches 908; Indels 865; Gaps 127;

QY 1 MNHUYKVFKNKATGTMVAEACAKSHSGSSSTAGOVGSS-----P 42
Db 1 MNKIFNVIMVMTQTWVWVSELRTHTKRASATVETAVLATLLFATVQANATDEDELDP 60
QY 43 VIRLTRVATLAILVIGATLNGSAYAQNNKSIATGTTCCNDNDSASNEASIAIGSLAKAHA 102
Db 61 VVR-----TAPVLSFHSKDEGTEKEVTENSNGYIFDNKGVKLAGAITLUKAGDNLKIKQ 115
QY 103 NOAIAIGSKPDPNRAANAKAGSHARK-----ESIAIGGDVLAEGD-ASTAIG 151
Db 116 N-----TDESTNASSFTYSILKDLTDITSVATEKLSFG-----ANGDKVDITSD 159
QY 152 SDDLYLDNRNSTNSKYPNGLLSTL---IQNHIVL-----ROI RDSNGS 190
Db 160 ANGLKLAKTGNVHNLGLDSTLPDVAVTNTGVLSSSSFTPNVDEKTRAAATVKOVNLNAGWN 219
QY 191 QKYRRTAAEGHASTAVCAMAYAKGHANAFGTRSTAGNYS LAVGLTAKAEKCYTTAIGS 250
Db 220 IKGAKTAG-GNVESVDLVSAINNVEFIT--GDKNT-----LDVVLTA-K-ENKKTVEVKF 269

QY 251 NAAQINYGALALGADTRVDLDYGIATGYGSQLNNNNNNNNNNKAYVPEGNGSNIKSSKAT- 309
Db 270 TPK-----TSV-----IKKDGKLFCKENNDTNK-----VTSNTATD 302
QY 310 ----GNGLFSIGSSTTKRKIIINVAGAGYEDTDVAVNAQLKAVENLAKRQ----- 353
Db 303 NTDEGNGLV-----TAKAVI-----DAVNKAGWRVYKTTTANGQNGDFATVASGTN 347
QY 354 ITPFKG-----DNGGTGV-----KKLGET--LTIKG----- 378
Db 348 VTESGDGTTASTVKTONGNGITVYKDYDAKVGDLKFPDSDRKIADVTTALTAVTGKVAEIA 407
QY 379 -ETQADKLTDDNNIGVYVVDNNTGLVKVLAKNLSG-LETVS--TKNLTAASEKVTYVSGNNTA 435
Db 408 KEDDKKKLVNAGDL-VTALGNLSWKAKEADTDGALBSISKDQEVKAGETVTFKAGNKLK 466
QY 436 ELQSG-----GLTFTPTTNASTD-KTVYGTGDLKFT--DNSNTALEDTTR 477
Db 467 VKODGANFTYSLQDALTGLTSITLGGTTNGNDAKTVINKDGLTITPAGNGGTTGTNTIS 526
QY 478 ITRDKIGFSNKA-----GTVDENKPYLDKD 502
Db 527 VTKDGIKAGNKAITNVASGLRAYDDANFDVLNNSATDLNRHVEDAYKGLLNLNKNANKQ 586
QY 503 KLVKGNSTLNNGL-----TVNNTIGGSN-KQ----- 528
Db 587 PLVTDSTAATVGDRLKLGWVSVTKNGTKBESNOVKQADEVLFTCAGAATVTSKSENGKHT 646
QY 529 -----IQVGADGIKPADVNVVNSNAKFT--TRITE-----EELGFADAD- 567
Db 647 ITVSVAEKTRADCLEKDGDTIKLVNDQNTDNVLTGVNNGTAVTKGFFETVKTGATDADR 706
QY 568 GKVDKSPYLDKKQLQVGVKITKDSGINAGDOKISNVKD-ATDSTDVAVYKOLKQVQDD 626
Db 707 GKV-----TVKDATANDAKKAVTVKOVATFATNAINSAATFVKTENUTTS 748
QY 627 ADGALOSFIRDEKQGEFTISNLSNG--NTPNTFITF-AGEN-----GISISNDIA 677
Db 749 IDE-----DNPTDNGKDDALKAGDGLTFKAGKNLKVKROGKNTFDLA 791
QY 678 KG-KVKV-----GIDPINGLT--TPKLTVGSDDGKTQLVIEOVASGNDTKNI-IR 724
Db 792 KNLEVKTAKYSDTLTIGGNTPTGGTTATPKVNTISTADGLN--FAKETADASGSKNYLK 849
QY 725 GLSPTL--PSITNAGGVRTTEQGNITISDEKSKAAASIGDILNTGPNLKNNSVGFVST 782
Db 850 GIATLTLEPS-----AGAKSSHVDLNVDATK-KSNAASIEDVLRAGWNIQNGNNDVYAT 904
QY 783 YNTVDFIDGNATTAKVTYDETNOTSKVTYDVNVDEKTIELTDGNGKTNKIGVKTTLTTLTT 842
Db 905 YDIVNFTDDSTGTTTVT--TQKADGKADVKIGAKTGVIKDHNGK-----LFTG 952
QY 843 NANGKATNFTTND-----ALVNAKDAENLNTLAKEIHTTKGTADTALQTFVKVKD 895
Db 953 KDLKDANNGATVSEDDGKTGTGLVTAKTVIDAVNKGSRVTEGATAET----- 1002
QY 896 GAT-----DDETLTVGKDGTVNGKTVNTLKGNGTLVATNKDG---TVTFGINTOSGL 947
Db 1003 GATAVNAGNAETVT-----SGTSVN--FKNGNATTATVSKDNGNINVDYVNGDGL 1052
QY 948 KAG-----DSTTLNKKDGLSINKPA-----SNEQIOVGADGVK-----FAKVDK-- 985
Db 1053 KIGDKKIKIVADTTTLTVTGKGVSPACANSVNNKVLVNAEGLATALNLSWTAKADKYA 1112
QY 986 -----GNSSTGIDGTSR----- 997
Db 1113 DGESEGETDOEVKAGKVTFFKAGKNLKVQKSEKDFTYSLQDTLTGLTSITLGGTANGRND 1172
QY 998 ----ITKDQIGFTGANGSLDTPKHLTKDKLVKGEVEVTNTGINAGCKKLTNIOGSDITQ 1053
Db 1173 TGTVINKDGLTITLUGAAAGT-----DASNGNTISVTKDGISAGNKKEITNVKSA---- 1222

QY	1054	NSNDAYTGRVYDLKTELESKINSAAKTAQNSLHEFSVADEQGNHFTVSNPYSSYDTSKT	1113
Db	1223	-----LKYKDTQ--NTADETQDEPHA-----AVKN	1247
QY	1114	SDVITFAGBNGITTKV---NRGVVRVIGIDQTK-----GLTTP-----KLJVGNNGKGI	1160
Db	1248	ANEVEFGKNGATVSAKTDNNGKHITVTIDVAEAKVGDGLEKDDTKLKLKLVNDTNGNLL	1307
QY	1161	-IDSKDQNTITGLSN-----TLANVTN-DGAGHALSOG-----LANDTKTRAASICDV	1208
Db	1308	TVDATKGASVAKGEFNAVTTDAATAGGTANMERKVVVKGSGNGATATETDKKKVATVDV	1367
QY	1209	LNAGFNLOGNEAVFVSFTYDTRIDGNATTAKVTYDDT-----SKTSKVYVYVNDNK	1263
Db	1368	AKA-----INDAATFVKVENDSDATIDDSPTDGDANDALKAGD	1405
QY	1264	TIEVTSKKLVKVTTLTKTSANGNATRFSAADGALVKASDIAPHNLNLAGDIQTAKGA	1323
Db	1406	TLTLKAGKNLKVK-----RDGNITITALAN-DLSVKSATVSKLS-----	1444
QY	1324	SOASSASVVDADCNKVIYDSDKKYQVNDKGOVDKNEKAKDLVAQAOTPDGTLAQM	1383
Db	1445	-----LGTNGKNVITSDTKLGNFAKD-----SKTGDANIHLNGIATLTDTLNS	1491
QY	1384	NVKSVINKEOVNDANKQGINEDNAP-----IKGLENAKDTTKNAAVTVGDLNAVAQT	1438
Db	1492	GATTNLGGNGITDNEKKRAASVDVLNAGNVNKGVPASANNQVENI-----DFVATYDT	1546
QY	1439	PLTFAGDTGTAKKLGTLTIKGGQDITDKNLTDNNITG-----VWAGPDG--FTVKLAKDLT	1492
Db	1547	VDFVGBKDDT-----SVTVE--SKDNKRTVEKIGAKTSVIKDHNKLGFTGKELKQAN	1598
QY	1493	NLNSVNAAGTRIDKGTISFVDANGQAKANTPULSANGCLDL---GKRISNIGAAYDDNDA	1549
Db	1599	N-----NGVTVTETDCKDEGNLGLTAKAVIDAVNAGWRVKTGTGANGONDD--	1644
QY	1550	VNFQFNEVAKTVNVLNNSQSGASLPFVVTYDANGKPINGTDGPKQAKAIGA-DGKYIHA	1608
Db	1645	-----FATVAGSTN-----VTFADG---NGTTAEVTKANDGSIIVKYNVK	1681
QY	1609	NANGVPVDKDKPITTDADKLANLAHOK-----PLDAGHOVVASLGGNSDAITLTIKSLTP	1665
Db	1682	VADGLKLDGD-KIVADTITLT--VADGKVTPANNNGDKFVDASGLADALNKLSWTAT--	1736
QY	1666	QIDTPTGNANAGQAQSLPSLSAAQSNASVSKDVLNVGNFLNTNHNQVDF-----VK	1718
Db	1737	-----AKEGTGEVD--PANSAGQEVKAGD-KVTERAGDNLKIKQSGKDFYSLAKKELK	1787
QY	1719	AYDTVNFVN---GTGADITSVRSADGTSNITV---NTALAATDDGDNVLKAKDCKFYK	1772
Db	1788	DLTSVEFKDANGGTGSESTKI-----TKDGLTITPANGAGAAGANTANTISVTKDG---	1838
QY	1773	ADDLPNGLSLKAGKSASDAKPTGLSLVNPAGKSGTGDVALANLNSKAVFKS-----K	1826
Db	1839	-----ISAGNKA-----VTNVYVGLKKFGDHTLANGTVADFEKHVDNAYK	1879
QY	1827	DGHTTTTSSDGISIQCKDNSSITSLKDLNVG-----GKVIS---NVGKTKDTDANV	1878
Db	1880	DLT-----NLDEKGDADNPNTADNTAATVVDGLRGLGWISADKTTGEPNQBYN-AQV	1930
QY	1879	QQLNEVRNLLGLG-NAGNDNAGNOVNIAIDKDP-----NSGSSSNRTVIKAGTV	1928
Db	1931	RNANEVFKSGNGINVSQKTLNGLTRVITFEKAKGEVKSNEFTVKNAQDSETNLVKVGDM	1990
QY	1929	LGKGKGNND--TEKLTAGGVQGVQVDKNGANGDLNWWVKTKDGSKKALLATYNAAGQTN	1986
Db	1991	YYSKEDIDPATSKPMTG---KTEYKVVENGKV-----VSANGSKTEVTTLTKNGSG---	2037
QY	1987	YLTNNFAEALDRINEOCIRFFHVNDGNQBPVQGRNGIDSSASGKHSVAIFQAK-----	2041
Db	2038	YVTGN--QVADAIKAGSFEI-----GIADAAEAEKFAESAKDKQLSKD	2079
QY	2042	-----ADGEAAVAIGRQTAGNOSTAIGDNAQATGDQSIAGTGNVTVTKHSIGAIQDPS	2095

RESULT 3
US-08-913-942-4
Sequence 4, Application US/08913942
Patent No. 6200578
GENERAL INFORMATION:
APPLICANT: St. Gene, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 398-3249
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

Thu Sep 13 14:18:09 2001

REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

Query Match 7.2% Score 841; DB 1; Length 1912;
Best Local Similarity 23.0%; Pred No. 1,6e-38;
Matches 475; Conservative 255; Mismatches 657; Indels 676; Gaps 99;

QY 1773 ADDLMPNGLKAGKASDAKTTPTGLSLVNPAGKSGTGDVAVALNNLSKAVFKS-----K 1826
DB 1839 -----ISAGNKA-----VTNVVSGLKKFGDGHITLANGIVAFERKHYDNAYK 1879
QY 1827 DCTTTTSSDGSISQKDNSSITLSKDLNVG-----GKVIS---NVGKGTKDTDAANY 1878
DB 1880 DLT-----NLDEKADNPVTADNTAATVGDRLGLGWISADKTTGEPNOEYN-AQV 1930
QY 1879 QQLNEVRNLLGIG-NAGNDNADGNQVNTADIKDP-----NSGSSNRRTVIKAGTV 1928
DB 1931 RNANEVKEKSGNINVSCKTLNTRVITFELAKGEVVKNSNEFTVKNADGSETNLVKVGDGM 1990
QY 1929 LGGKGNND--TEKLTGGVGVGVGDGNAGDLSNVVVKTKQDGSKKALLATYNAAGQTN 1986
DB 1991 YSKEDIDPATSKPMTG-----KTEYKVENGV-----VSANGSKTEVTLTNKSGS--- 2037
QY 1987 YLTNPAEAIADRIEIOGIRFFHVNDGNQBPVVQVGRNGIDSSASGKHSVAIGFOAK----- 2041
DB 2038 YVTGN--QVADALAKSGFEL-----GLADAAEAKEAFESAADKQKLSKD 2079
QY 2042 -----ADGEAAVAIGROQAGNOSTAIGDNAQATGDQDQSIAGTGNVVTGKHSGAIGDPS 2095
DB 2080 KAETVNAHDKVRPANGLNKTV---SAATVESTDANGDK-----VIT-----T 2118
QY 2096 TVKADNSYSYGNNOQFIDATQTDVFGV---GNNITVT-----ESNSVALGNSAISAG- 2145
DB 2119 FVKTD-----VELPLTOIYNTDANGNKIVKKADGKWYELNADGTASNKEVTLGN 2167
QY 2146 --THAGTQAKSDGTACTTTTACATGTVKGFAGOTAVGAVSVG-----ASGAE 2191
DB 2168 VDANGKKVVKVTENGADKWTYTNADGAADTKGEVSNKVDSTDEKHVRLDPNNSNGK 2227
QY 2192 RRIQNAAGEVSATSTDAVNGSOLYKATOGIAN---ATNELDHRIHQENKANAGISSAM 2248
DB 2228 VVIDNVANGELISATSTDAINGSOLYAVAKGVTNLAGOVNLEKGVKVKGRADAGTASAL 2287
QY 2249 AMASMPQAYIPGRSMVTTGGTATHNGOGAVAGLSKLSDNCGWVFEKINGSADTQGHVGAIV 2308
DB 2288 AASQLPQATMPGKSMVAIAGSYYGQGLAIGVSRISDNGKVIIRLSGTTNSQKGTGVA 2347
QY 2309 GAGPHF 2314
DB 2348 GVGYQW 2353

RESULT 4
US-08-409-995-4
Sequence 4, Application US/08409995
Patent No. 5646259
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
CORRESPONDENCE ADDRESS:
ADDRESS: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.

QY 1 MNHYIKVIFPKATGTFMAVAECAKSHSG-----SSSSTAGQVGVSSPV 43
DB 1 MNKIFVNVNVTQTWVWVSELTRTHTKRLNRNGDPVLATLLFATVQANATDEDELDPV 60
QY 44 IRLTRVATL-----AILVIGA-----TL 61
DB 61 VRTAPVLSFSDKEGTGEKEVTENSNGIYFDKNKGVLKAGATLTKACDNLKXQXTDEX 120
QY 62 NSGAYQON-----NSKIAFGTTGNN-DNASASNEASIAIGSLAKAHANQIAI 108
DB 121 NASFTYSLKKDLTDLTSVATEKLSFGANGDKVDLTSDANGKLAKTNGNVHLN---GL 177
QY 109 GGSKPD-----PRQANOK-----AGSHAKGESIAIGGOVLAEGDA 146
DB 178 DSTLPDAVNTGVLSSTSPFPNOVEKTRAATVQKVDLNGWNIKAKTI--AGNV----- 229
QY 147 SIAIGSDDLYLDRNSTSKYPNGLLSTLIQNHVTLROI RDSNGSQKYRRTAAEGHASTAV 206
DB 230 -----ESVDLVSAYN--NVEFITGDKNTL---DVLTAKENXKTEVKFPTKSVIKE 279
QY 207 GAMAYAKGHFANAGTRSTAEGNYSVAVGL-TAKA-----EKGYTI-AIGSNAQAALNYG 258
DB 280 GKLFTEKENNDNTKVTSTATDNTDEGNGLVTAKAVIDAVNKAGWRVKTATTANGCQDPA 339
QY 259 ALALGADTRVDLDYGIAGLYGSQLNNNNNN--NNKAYVPEGNSNIKSSK----- 307
DB 340 TVASG--INVTFESGD--GTTASVTKDTNGNGITVKYDAKVGDLGKFDSDKKIYADTTAL 395
QY 308 -ATGNGLFSIGSSTTKRLIINV-----AGYEDTDAVNVNAQLKAVENLAKRQ 353
DB 396 TVTGGKVAEIAKEDDDKKLVNAGDLVTALGNLSWKAKAEADTDG-----ALEGISKQ 448
QY 354 ITFKGDDNGTGVKKLIGETLTIKGE-----TQADKLTDNNNIGVVTDDNTG 400
DB 449 -----EVRKAGEITVFKACKNLKVKQDGANFTVSLQDALTGLTSLTGOTTNGG 496
QY 401 LKVKLAKNLGSLG-----TVSTKNLTASEKVTVVGSNNNTAELOSGGL-----TFTP 446
DB 497 NDAKTVINKDGLTITPAGNGGTGTNTISVT-KDGLKAGNKAITNVASGLRAYDDANEDV 555
QY 447 TTNASTDKTVYGTGDKLKTDNNS-----TALEDITRITKD---KIGF--SNKAGTVDE 494
DB 556 LNNSATDLNRHVEDAYKGLLNLEKNANKQPLVTDSTAATVGLRLKLGWVSTKNGTKEE 615
QY 495 NKPYLDKDKL---KVGNSLT-----NNGGLTVNNTIGGSNKQIOVGDG--IKFADVNVV 545
DB 616 SNQVQADEVLFTGAGAAVTVTSKSENGKHTITVSVAEKADCGLEKDGOTIKLKVQNTQT 675
QY 546 SNAKECT--TRITE-----EEIGFADAD-GKVDKSPYLDKKQLQVGVKTKTKSGINA 597
DB 676 DNVLTVGNNGTAVTKGFEVTKGTATDADRGV-----TVKDATAND 717
QY 598 GDQKISNVKD-ATDDTDAVTVYKQIKOVQDADGALQSFSIRDEKGOEFTISLNSYNG--N 654

Db 718 ADKKVATVKDVATAINSAATFVKTEMLTTSIDE-----DNPTDNGKDD 760
QY 655 TPNTFETITF-AGEN-----GISISNDIAG-KVKV-----GIDPINGLT-TPK 695
Db 761 ALKAGDTLTFKAGKLNKVKRDGNITFDLAKNLEVKTAQVSDTLTIGGNTPTGTTATPK 820
QY 696 LTVGSDKDKTQLVIOVASGNDTKNI-IRGLSPFL--PSITNAGGVRTTEQONTTISDE 752
Db 821 VNITSTADGLN--FAKETADASGKNVYLKGIATTLTPE-----AGAKSSHVDLNDATK 874
QY 753 DSKAASIGDILNTGFLNKNNSGVSTYNTVDIDGNATTAKVTYDNTOTSKVTYD 812
Db 875 -KNAASIEDVLRACNWIQGNNGNVDYVATYDVNTDSTGTTTVV--TQADCKGAD 931
QY 813 VNYDEKTIELTDGDKNTKIGVKTTLTTTNANGKATNFSTTDND-----ALVNAKDI 865
Db 932 VKIGAKTSVTKDHNGK-----LFTGKDLKDANNGATVSEDDGKDTGTLVTAKT 981
QY 866 AENLNTLAKEIHTTKGTADTALQTFVKVKDGAT-----DDETITVKGDTONGKTVNTLK 920
Db 982 IDAVNKGSRVTVGEGATAET-----GATAVNAGNAETVT-----SGTSVN-- 1021
QY 921 LKGENGLTVATNKG--TVTFGINTOSGLKAG-----DSTLLNKDGLSIKNA-- 966
Db 1022 FKNGNATTATVSKONGNINVKYDVNVCDGLKIGDKKIIVADTTTLTVTGKVSVPAGANS 1081
QY 967 -SNEQIQVGADGVK-----PAKVDK----- 985
Db 1082 VNNKKLVNAEGLATALNLSWTAKADKYADGESEGTDOEVKAGDKVTFKAGKLNKVKQ 1141
QY 986 -----GNSSTGIDTSR-----IYKDIQIGTCANGSLDTPPHLTCKD 1023
Db 1142 SEKDFYSLQDTLTGLTSITLGTANGRNDTGTVINKDGTLITLANGAAGT-----DA 1195
QY 1024 LKVGVEITWTGNAGKKTNIQS-----GDITQ-----NSNDA-- 1058
Db 1196 SNGNTSVTKDGISAGKKEITNKVSALKYKQNTADETQDFEFAAANKNAMEVEFGK 1255
QY 1059 -----VTGGRVYD-LKTELESKIN-----SAAKTAQ 1083
Db 1256 NGATVSAKTDNNGKHTVTIDVAEAKVGDLGKDTGDKLKLVKVDNTDGNLLTVDATKGAS 1315
QY 1084 NSLHESV-----ADQGHFTV-SNPYSYDTSK-----TSDVIT 1118
Db 1316 VARGEFNAVTTDATTAGTNNANERGVVVGKSGNGATATETDKKKVATGVGVAKAINDAAT 1375
QY 1119 FA-----GENG-----ITTKVNGK--VVRVGDIDQTKGL----- 1144
Db 1376 FVKVENDDSATIDDSPTDDGANDALKAXDTLTLKAGKLNKVKRDGNITFALANDLSVKS 1435
QY 1145 --TTPKLTGVNNGK-----GIVDSKDGQNT-----ITGLSNTLAN-VTNDGAGHA 1188
Db 1436 ATVSDKLSLGTNGKNVNTSDTKGLNFAKDSKTGDDANIHLNGIASTLTDTLLNSGATTN 1495
QY 1189 LSOGLANDTUKTAASTGVDVNLNAGFNLOG-----NGEAVDFVSTYDVTDFIDGNATT 1240
Db 1496 LGNGITDNEKKAASVKVDVNLNAGVNRGVKSPASANNQVENIDFVATYDVTDFVSGDKDT 1555
QY 1241 AKVTYD--DTSKTSKVVYDVNVNDNKTIEVTSKDKLGVKTTTLTK-----TSANG 1287
Db 1556 TSVTVESKONGKTEV-----KIGAKTSVTKDHNGKLTGTGKELKDANN 1598
QY 1288 NATKFAAD-----GDALVRASDIATHLNTLAGDIOTAKGASQ-----ASSASYVDADG 1337
Db 1599 NGVTVTETDCKDEGNGLVTAKAVIDAVNAGWRVKTGTGANGQNDPFATVAGSGTNTVTFADG 1658
QY 1338 NKVIYDSTDKKYYQVNDKG-QVDKNEVAK-----DKLVAQA--QTPDGTLAQNNVKS 1387
Db 1659 N-----GTTAEVTKANDGSITVYKYNKVAADGLKLDGDKIVADTVTLTVADGKVTAPN-- 1710
QY 1388 VINKEQVNDANKOGINEONAFKIGLENAKDTKTKNAAVTVGDLNNAVAQTPLTFAAGDTG 1447
Db 1711 -----NGXGKK---FXDASGLAGCLNKLXSTATACKECT-GEVD-----PANSAGQ-- 1752

QY 1448 TTAKKLGELTITKGGQDTNKLTDNNIGVVVAGTGDGFTVKLAKDLTNLSV-----NAG-- 1500
Db 1753 --EVKAGDKVTFRAG-----DNLKIKQSKKDFYSLKKELKDLTSVEFKDANGGTG 1801
QY 1501 --CTRDEKGISFVDANGQA-----KANTPVLSSANGDLIDGKKRISNI-----GAAV 1544
Db 1802 SESTKTKDKGLTTPANGAGAAGANTANTISVTKDGISAGNAKAVTNVYVGLKKRFGDGHLL 1861
QY 1545 DDNDVNF-KOFNEVAKTVNNLN 1566
Db 1862 ANGTVADFEKHYDNAYXDLTNLD 1884

RESULT 5

US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-685-467-4

Query Match 7.2%; Score 841; DB 3; Length 1912;
Best Local Similarity 23.0%; Pred. No. 1.6e-38;
Matches 475; Conservative 255; Mismatches 657; Indels 676; Gaps 99;

QY 1 MNHIIYKVIENKATGTMAVAECAKSHSGG-----SSSSTAGQVSSPV 43
Db 1 MNKIFNVIMVMTQTWVWVSELTRTHTKRLNRGDPVLATLLPATVQANATDEDELDPV 60
QY 44 IRLTRVATL-----AIIIVICA-----TL 61
Db 61 VRTAPVLSFHSDEKTEGTEKEVTENSNWGIYFDNKGVLKAGAITLKAGDNLKXKQTXDXT 120
QY 62 NGSAYAQN-----NSKIAFGTTGNN-DNASASNEASIAIGSLAKAHANQAI 108

Db 121 NASSFTYSLKDLTDLTSVATEKLSFGANGKVDLITSDANGLAKTNGVHLN---GL 177
QY 109 GSKPD-----PRNOAOK-----AGSHAKGESTAIGDVLAEODA 146
Db 178 DSTLPDAVNTGVLSSSFTPNDEKTRAATVKVNLNAGWNIKAKT---AGGNV----- 229
QY 147 STAIGSDLLYDRNSTNSKYPNGLSLTLQNHITVLRQJRDNSGOKYRRTAAEGHASTAV 206
Db 230 -----ESVDLVSAYN--NVEFITGDKNTL---DVLITAKENKKTTEVKFTPKTSVIKEKD 279
QY 207 GAMAYAKAHFAMAFGRSTAGNYSVLGL-TAKA-----EKGYTI--AIGSNAQAINYG 258
Db 280 GKLFTGKENNDNTNKVTSNTATNDDEGNLVTAKAVIDAVNKAAGRWKVTITANGQNGDEFA 339
QY 259 ALAAGADTRVLDYDIALGYSGOILNNNN--NNKAVVPEGNSNIKSSK----- 307
Db 340 TVASG--TNVTPESGD--GTTASVTKDTNGNITVKYDAKVGDLGLKFDSDKKIIVADTTAL 395
QY 308 -ATGNLFSIGSSTIKRKLIING-----AGYEDTDAVNAQLKAVENLAKRQ 353
Db 396 TVTGKVAEIAKEDDKKLVNAGDLVLTALGNLSWKAKAEADTDG-----ALEGISKQ 448
QY 354 ITFKGDNGTGKVKLGETLTIKGE-----TOADKLTDNNNIGVVTDNNTG 400
Db 449 -----EVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTLGLTSITLGGTTNGG 496
QY 401 LKVKLAKNLGLE-----TVSTKNLTASPKVTGSGNNTAELOSGGL-----TTP 446
Db 497 NDAKTVINKDGLTIIPAGNGTGTINTISVT--KDGKAGNKAITVWASGLRAYDANDFV 555
QY 447 TTNASTDKTVYDGLKFTDNSN-----TALEDTRITKD---KIGF--SNKAGTVDE 494
Db 556 LNSATDLNRHVEDAYKGLLNLNEKNANKOPLVTDSTAATVGDRLKGLVWVSTKNKTREE 615
QY 495 NKPYLDKDL---KVGNSL---NNGGLTVNNTIGGSKQLOVGADG--IKFADVNVV 545
Db 616 SNOVKQADEVLFTGAGATVTSKSENGKHTITVSAETKADCGLEKDGDTIKKYVDNONT 675
QY 546 SNAAPGT--TRITE-----EEIGPADAD-GKVDKPSYLDKQLOVGQVKTIKDSGNA 597
Db 676 DNVLTVGNNGTAVTKGFFETVKTATDADRKV-----TVKDATAND 717
QY 598 GQOKISNVKD-ATDDTDVATYKQLOVQODADGALQSFSIRDEKGOEFTISNLYSG--N 654
Db 718 ADKAVTAKDVATAINSAATFKVTENLTTSIDE-----DNPTDNGKDD 760
QY 655 TPNTFETITF-AGEN-----GISISNDIAKG-KVKV-----GIDPINGLT-TPK 695
Db 761 ALKAGDTLTFKAGKNLKVRDGNITFDLAKNLEKVTAKVSDTLTIGGNTPTGGTTATPK 820
QY 696 LTVGSDKDGKTQLVIBQVASGNDTKNI--IRGLSPTL--PSITNAGGVRTTEGNTITSDE 752
Db 821 VNITSTADGLN--FAKETADASGKNVLYKIATLTLEPS-----AGAKSSHVDLNVDAIK 874
QY 753 DKSKAASIGDILNTGNLKNNSVGFVSTYNTVDFIDGNATKATVYDETQTSKVTYD 812
Db 875 -KSNAAISIEDVLKAGNIQNGNVDYATVYDVTNFTDDSTGTITVIV--TQKADCKGAD 931
QY 813 VNYDEKTIELTDGNGKTNIGVKRTTLTTTNNANGKATNFSTTDDND-----ALVNAKDI 865
Db 932 VKTGAKTSVYKDHNGK-----LFTGKDLKDANNGATVSEDDGKDTGLTAKTV 981
QY 866 AENLNTLAKEIHTTKGTADTALQTFKVKKDGAT-----DDEITVKGKDTQNGKVTNLUK 920
Db 982 IDAVNKSGRVVTGEGATAET-----GATAVNAGNAETV-----SGTSYN--- 1021
QY 921 LKGENGLTVATNKG--TVTFGINTQSLKAG-----DSTTLNKDGLSLKPNPA--- 966
Db 1022 FKGNATTATVSKDGNINVKYDVNVGDLKIGDKKIIIVADTTTLTVTGKVSVPAGANS 1081
QY 967 -SNEQIQVGADGVK-----FAKVDK----- 985

Db 1082 VNNKKLVNAEGLATALNLSWTAADKADYADGESEGETDOEVKAGDKVTFKAGKNLKVQ 1141
QY 986 -----GNSSTGIDGTSR-----ITKQDQIGFTGANGSLDTTKPHLTKDK 1023
Db 1142 SEKDFYTSLODTLGLTSITLGGTANGRNDTGIVINKDGLTITLANCAAACT-----DA 1195
QY 1024 LKVEVEITNTGINAGGKKTNIQS-----GDITQ-----NSNDA----- 1058
Db 1196 SNGNTISVTKDGISAGNKEITNVKSALKTYKDTQNTADETQDKAEFHAHVKNANEVEVGK 1255
QY 1059 -----VTGGRVYD-LKTELESKIN-----SAAKTAQ 1083
Db 1256 NGATVSAKTDNNGKHVTIDVAAEAKVGDLGKLEKTKLKVNDTNGNLLTVDATKAS 1315
QY 1084 NSLHEFSV-----ADEQCNHFTV--SNPYSYDTSK-----TSDVIT 1118
Db 1316 VAKGEFNATVDTAATAGTNNANERGVVVKSGNGATATETDKKVVATVDVAKAINDAAT 1375
QY 1119 FA-----GENG-----ITTKVNGK--VVRVGIQDTKGL----- 1144
Db 1376 FVKVENDDSATIDDSPTDDGANDALKAXDITLTKAGKNLKYKRDGKNITFALANDLSVKS 1435
QY 1145 --TTPKLTVGNNGCK-----GIVIDSKDGONT---ITGLSNTLAN-VTNDGAGHA 1188
Db 1436 ATVSDKLSLGTNGKNVNTSDTKGLNFAKDSKTGDANIHLNGIASLTDLTLLNSGATTN 1495
QY 1189 LSOGLANDTDKTRAASIGDVLNAGFNLOG-----NGEAVDFVSTYVDVDFIDGNATT 1240
Db 1496 LGGNGITDNEKKXAAASKVDVLNAGWNRGVKPASANNQOVENIDFVATVTDVDFVSGDKDT 1555
QY 1241 AKVYID--DTSKTSKVYVDVNVNDKTIEVTSKKLGVKTTTLTK-----TSANG 1287
Db 1556 TSVTVESKDNCKRTEV-----KIGAKTSVIKDHNGKLTGKELKDDANN 1598
QY 1288 NATKFSAAD---GDALVKASDIATHLNTLAGDIQTAKGASQ-----ASSASYSVDADG 1337
Db 1599 NGVTVETDGRDEGNLVTAKAVIDAVNKAAGRWKVTTCGANGQNDDEFATVASGTVVTFADG 1658
QY 1338 NKVIYDSTDKYQVNDKG-QVDKNKEVAK-----DKLVAQA---QTPDGTLAQMNVKS 1387
Db 1659 N-----GTTAEVTKANDGSITVKYNVKVADGLKLDGDKIVADTTVLIVVADGKVTAPN--- 1710
QY 1388 VINKEQVNDANKKQGINEDNNAFIKGLENAADTKTKNAAVTVGDLNNAVAQTPLTFAAGDTG 1447
Db 1711 -----NGXCKK---FXDASGLACLNKLSXTATAGKEGT-GEVD-----PANSAGO-- 1752
QY 1448 TTAKKLGETLTIKGQDTNKLTDNNIGVAVAGTGFVVKLAKDLTNLSV-----NAG--- 1500
Db 1753 --EVKAGDKVTFKAG-----DNLKIKOSXKDTYSLLKELKDLTSVEFKDANGGTG 1801
QY 1501 --GTRIDEKGISFVDANGOA-----KANTPVLSANGLDGGKRISNI-----GAAV 1544
Db 1802 SESTKITKDGTLTITPANGAAGANTANTISVTKDGISAGNKAVTNVVSGLKKFGDGHTL 1861
QY 1545 DDNDAVNF-KOFNEVAKTVNNLN 1566
Db 1862 ANGTVADEFKHYDNAYXDLITNLD 1884
RESULT 6
US-08-409-995-2
; Sequence 2, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685.467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-685-467-2

Query Match 5.68; Score 659.5; DB 3; Length 1098;
Best Local Similarity 22.9%; Pred. No. 8.5e-29;
Matches 323; Conservative 168; Mismatches 457; Indels 461; Gaps 55;

QY 1008 ANGSLDTTKPHLTBKDKLV---GEVEINTGINAGGKKTINIQSGDITON----- 1054
DB 49 ANNTPTVT-----NKLKAYGDANFNFNNSIADAEKQVQKAYKGLLNLEKNASDKLLV 102
QY 1055 -SNDVAVTGRVYDLKTELESKINSAKTAQNSLHSEFVADEQNHFTVSNPYSDYDTKT 1113
DB 103 EDNTAATVGNLRKLGWLSSKNGTRNEKSOQVKHAEVLFEKGKGVQVT---STSENGKH 159
QY 1114 SDVITFAGENGITTKVNGVVRVIGDQTKG-LTTPKLV-GNNCKGIVIDS--KDGQNT 1169
DB 160 TITFALAKDLGVKTATVSDTLTIGGGAAGATTTPKVNVTSTDTGLKFAKDAAGAGDGT 219
QY 1170 I--TGLSNTLANVNDGAGHALSOGLANDTDKTRAASIGDVLNAGFNLO-----GNG 1219
DB 220 VHLNGIGSTLTD-TLVGSPATHIDGGDQSTHYTRAASIKDVLNAGWNKIKGVKAGSTTGOS 278
QY 1220 EAVDFVSYDVTDFIDGNATTAKYVYDDTSKTSKVYDVVNDKNKTIETVSOKKLGKVT 1279
DB 279 ENVDVFTYDVTVEFLSADTFETTTVTVDSEKNGKRTVEKIGAKTSVIKEDGKLF 338
QY 1280 LTKTSANGNATKFSAADGDALYKASDI-----ATHLNTLAGDIQT-AKGAQA 1326
DB 339 ETNKVDGANATE-DADEGKGLVATKDVIDAVNKTGWRIKTTDANGQNGDFATVASGTNVT 397
QY 1327 SSSASYDA-----DGNKVYDYDTSKKYQVNDKQVQDNKKEVAKDLVAQAQPTDGT 1380
DB 398 FASGNGTTATVNTGDTGVKYDA-----KVGDLGLKD-----GDKIAA-----DTTA 440
QY 1381 AQMNKSVINKQVNDANKKQGINEDNAFIKGLENAADTK---TKNAAVTVGDNLNVAQ 1437
DB 441 LTVN-----DGNKANNPKGVADVADEKLLVAKGLVTA--LNSLSW 482
QY 1438 TPLTFAGDTGT-----TAKKLGELTIKGGOTDNTKLTDDNIGVWAGTDGFTVKLAKD 1490
DB 483 TTTAAEADGGLDGNASEQEVKAGKRVFKAGK-----NLVKQEGANFTYSLQDA 533
QY 1491 LFNLSV-----NAGGTRIDPKGISFVDANGQA--KANTPVLSANGLDLOGKRISNICA 1542
DB 534 LFGLSITLGTGNGNAKTEINKDGLTITPANGAGANNANTISVTKDGISAGSQSVKNVVS 593

QY 1543 AVDDNDVAVNFKQFNEVAKTVNNLNQSNSSGASLSPFVVTDANGKPINGTDGKPKQAIKGAD 1602
DB 594 GLKKFGDAN---FDPLTSSADNLTKQND-----AYKG-- 623
QY 1603 GKYIHANANGVPVDKDKPITDADKLANLAHCKPLDACHQVVASLGGNSDAITLTNKS 1662
DB 624 -----LTNLDE-----KGTDKQTPVAD----- 641
QY 1663 TLPQIDTPTNTGNANQAQSLPSLSAAQSNAAQSVKQVLNVGNLQTNHNVDFEYKAYDT 1722
DB 642 -----NTA-ATVGDRLGLGWISADKTTGGSYE-----YHQ-----VRNAME 678
QY 1723 VNFYNGTGADITSVRSADGTSNITVNTALAAATDDGNNVLIKAKDG-----KPYK 1772
DB 679 VKFKSGNGINV-SGKTVMG-RREITFELAKGEVWKSNEFTVKTNGKETSLVKVGDKVYS 736
QY 1773 ADDLMPNGSLKAGKASDAKPTGLSLVNPNAKKGSTGDAVALNLSKAVPKSKDGT 1832
DB 737 KEDI-----DLTGQPKLKGNTVA 756
QY 1833 TVSSDGISIQCKDNSSITLSKDLNVGKVISNKGKTKDIDAANVQOLNEVRNLLGLCN 1892
DB 757 AKYQD-----KGCKVVS-----VTDNTEAT-----ITN 779
QY 1893 AGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGKGGNNDTEKLTATGQVGVVDK 1952
DB 780 KGSYVVTGNQV-----ADAIKSGFELGLADE 806
QY 1953 GNANGDLSNVVVKTKDGSKKALLATYNAAGOTNYLTNNPABAIIDRIINQGIREFHVNDG 2012
DB 807 ADA-----KRAFDDKTKAL-----SAGTTEIV-----NAHDK-----VRF----- 836
QY 2013 NOEPVVGRRNGIDSSASGKHSVAIGFQ-----AKADGEAAVAIGRTOAGNOSI--AIGD 2065
DB 837 -----ANGLNTKVSATVESTDANGDKVTTTFTVTDVPLTQIYNTDANGKKITKVVD 891
QY 2066 -----NAQATGQDSTAIGTGNVVTGKHSAGIDPSTVTKADNSYSVGNNOFIDATQT 2117
DB 892 GQTKWYELNADGTADMTKEVTILGNVDS-----DGKKVVKDND----- 928
QY 2118 DVFGVGNITVTESNSVALGNSALSAGTHAGTQAKKSDGTAGTTTATAGCTVKGFAQ 2177
DB 929 -----GKWHVHA-----KADGTADKT-----KGEVSN 949
QY 2178 TAVG-----AVSVG-----ASCAERRIONVAAGEVATSTDAVNGSOLYKATOGIAN---A 2225
DB 950 DKVSTDEKHHVYSLDPNDQSKGVVIDNVANGDISATSTDAINGSOLYAVAKGVTNLAGQ 1009
QY 2226 TNELDRHITHONENKANAGISSNMANAMASHPOAYIPGSRMVTGGIATHNGGAVAGLSKLS 2285
DB 1010 VNNLECKVNVKGRADAGTASALAAASQLPQATMPGKSMVATAGSSYQONGLAIGVSRIS 1069
QY 2286 DNGQWVFKINGSADTQGHVGAAGVAGFHF 2314
DB 1070 DNGKVIIRLSGTTNSQKTVGAAGVGYOW 1098

RESULT 8
US-09-377-155-32
; Sequence 32, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12

RESULT 10
US-08-913-942-15
; Sequence 15, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Fiehr Honbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:


```

; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/REF/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 679 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-913-942-15

Query Match 5.1%; Score 592; DB 4; Length 679;
Best Local Similarity 23.9%; Pred. No. 2.4e-25;
Matches 221; Conservative 121; Mismatches 285; Indels 298; Gaps 32;

QY 1433 NAAQTPLETFAGDTGTTAKLGLTL-----TIKGGQTDTKLTDNNIGVAGTD 1481
Db 10 NVVTQWVWVSELRTHTKCASATVAVAVLATLLSATVQANADENE----- 56

QY 1482 GFTVKLAKDLTNLSNAGGTTRIDEKISFVDANGAKANTPVLSANGDLGKRISNIG 1541
Db 57 -----DDEEELPEVQSVLRWSFK--SAKEGTGEQEGTTEVINLN-TDSSGNAVGSST 106

QY 1542 AAVDDNDVAFKQFN-----EVAKTYNNLNQNSGASLFFVTDANGKINGTDGPKQK 1596
Db 107 ITFKAGDKLNIKOSGNDFTYSLKELKNLTSVETELSF-----GANGNKVDIT- 155

QY 1597 AIKGADEKYYHANANGVPVDKGPITDADKLANLAHAKPLDAGHQVVASLGSNDAIT 1656
Db 156 -----SDANGLKLAKTNG-----NGQNSVH 177

QY 1657 LTNIKSLPQIDTPNTGNANAGQAQSLPSLSAAQSSNAASVKDVLNVGNLQTNHNOVDF 1716
Db 178 LNGIASTL-----TDLTAGTGTGHVDT--NIDAVNYHRAASVQDVLNSGWNITQGNVDF 231

QY 1717 VKAYDTYFVNGTGADITSVRSADGTSNITVNTALAA-----TDDGNVLIKAKGKFY 1771
Db 232 VRTYDVTDFVNGANANVSV--TADTAHKTTVRVDVTGLPQVYVTEDEGTVVKV-GNEYI 288

QY 1772 KADDLMPGSLKACKSASDAKTPGLSLVNPNAKGSTGDAVALNLSKAVEKSKDGT 1831
Db 289 KAKD-----DGSADNMOKVENGELAKTKVKLVASGT 320

QY 1832 TTVSSDGIISQKDNSSITLSKDLNNGVGGKVISNMGKTKDTPDAANVQQLNEVRNLLGLG 1891
Db 321 NPVK-----ISNADGTTEDTDAVSPKQLKALQD----- 348

QY 1892 NAGNDNADGNQVNIADIKDPNCGSSNRVRIRAGITVLGKGNNDTEKLTATGGVQGVGDK 1951
Db 349 -----KQVTL-----STSN-----AYANG----- 363

QY 1952 DGNANGDLSNWKTKQDKGSKALLATYNAAGOTNYLTNNPAEIDRINEQGITRF-PHVN 2010
Db 364 -----TDNDGK-----ATQTLN-----GLNFKFKSS 386

QY 2011 DGNQEPVVOGRNGIDSSASG-----KHSVAIGFOAKADGEAAVAIGRQTQAGNQSIA 2062
Db 387 DGEL-----LKISATGDTVTFTPKGSGVQVG-----DDGRASISKGANTTEG--LVE 431

QY 2063 IGDNAQTGDSQTAIGTGNVVTGKHSQA-----IGDPTVVRADNSYSV---GNNNQFI 2112
Db 432 ASELVESLNLKGMKVGVKGVSGELDGTSTKTLVKSGDKVTLLKAGDNLKVKQEGTN--FT 489

QY 2113 DAYQTDVFGVGNNTVTESNSVALGSNSAITSAGTHAGTQAKKSDGTAGTTTGTAGATCTVK 2172
Db 490 YALKDELGTGKVS-----VEFKDTANGANGASTKTKDKGLTTLTLANGANGATVTT----- 537

QY 2173 GFAGQTAVGAVSVGASGAERRIONVAAGEVSATSTDAVNGSOLYKATQGIAN---ATNEL 2229
Db 538 ---DADKIKVAGSGISAGNKAVNVAAGEISATSTDAINGSQLYAVAKGVTNLAGQVNNL 594

QY 2230 DHRHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNCOGAVAVCLSKLSNQG 2289
Db 595 EGVNKKVGRADAGTASALASQLPQATMPGKSMVSTAGSYOGQGLAIGVSRISDNKG 654

QY 2290 WFFKINGSADTQGHVGAAGVAGGFHF 2314
Db 655 VIIRLSGTTNSQKGTGVAAGVGQW 679

RESULT 11
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 4.8%; Score 565.5; DB 2; Length 1600;
Best Local Similarity 22.1%; Pred. No. 2.3e-23;
Matches 413; Conservative 242; Mismatches 709; Indels 505; Gaps 90;

QY 1 MNHIYKVFKNKATGTFMAVECAKSHSGSSSTAGOVSSPVRLTRVATLAILVTGAT 60
Db 1 MNKIRLKFSLKRNALVAVSELTR---GCDHST--EKGSEKPVV-TKVRHLALKPLSAI 53

QY 61 LNSAYAQ-NNSKIAFTCTTGNNDNASNEASIAIGSLAKAHANQAIAIGSKPDPNQA 119
Db 54 LLSLGMASIPQSVLASLQG-----MSVVHGTAATMQVDGNKTTIRN-- 94
```


QY 120 ANQKAGSHAKGESIAIGGDVLAEGDASTAIGSDLYLDRN-----STNSKYPNGLL 171
Db 95 -----SVNAIINWKFNIDQNEMEQFLOQESSNAFENRVT 129
QY 172 STLQHIVTLROIRDSNGSOKYRRTAAEGHASTAVGAMAYAKGFANAFG-TRSTAE-GN 229
Db 130 SDQISO---LKGILDSNG-OVELINP-----NGITIGKDAIINTNGEFTASTDLISN 176
QY 230 YSL-AVGLTAKAEKGYTTAIGSNAQAQINYGALAGADTRVDLDYG-----I 274
Db 177 ENIKARNFLEOTKOKAL-----AEIVNHGLITVGEKDSVNLIGKVKNEGVIYNGGSI 231
QY 275 ALGYGSOILNNNNNNNNKAY-----VPEGNSNITKSKATGNGLFSGSSTIKRK----- 324
Db 232 SLLAGOKITISDIINPTITYSIAAPENEAINEILGDIFAKG-GNINVRRAATIRNKKLSADS 290
QY 325 -----IINVAGYEDTDVAVNQAOLKAVENLAKRQITFKGDDNGTGKVKLGETLTI 375
Db 291 VSKOKSGNIVLSAKEGEAEIGGVISAQNOQAKG-GKLMIT-----GOKVTLTKTGAVIDL 343
QY 376 ---KGGETO--ADKLTDDNNIGVWTDNNTGLVKLAKNLISGLETVSTKNLTASER----- 425
Db 344 SKEGETYLGGERGEGKN-----GLOLAKTT-LEKGSTINVSKEKGGRAI 391
QY 426 ---VTVGSGNNTAE-----LQSGGLTFTPTTNASTDKTVYGGDGLKFTDNSNTALEDDT- 476
Db 392 VNGDIALIDGNINAOGSDIAKTGGFVETSGHDSIGDDVIDDAKEWLLDDPDVSIETLTS 451
QY 477 --RITKDKTGFNSKAGTVDE-----NKPYLKDKLK-----VGNSTLNNGGTVNNT 521
Db 452 GRNNTGENGYTGTGDKESPKNISKPTLTNSTLEQLILRGSVVNITANN-RIVVNSS 510
QY 522 IGGSNKQIOVGA--DGIK-----FADVNVNVSNAAKFTTTRITEE 560
Db 511 INLSNGLTLHTRKRGVKINGDITSNENGLTIKAGSWVDVHKNTITLGTGF-LNIVAGDS 569
QY 561 IGFADADGKVKKSPYLDKQLQVGVKITKD-----SGINAGDOKISNVKDAT 609
Db 570 VAF-EREG--DKARNATDAQITAQGTITVKNDDKQFRFNVSNLNGTGRGLKFIANNF 626
QY 610 DOTDA-----VITYKOLKOVQDADGALQSFSIRDEKGEFTISNLYSNG----- 653
Db 627 HRFDEGEINISGIVTINQTTKDKVYWNASKDSYWNVSSLTNTVQKFTPIKFVDSCSNQ 686
QY 654 ---NTPTFTTTFAGENGISINDIAKGVKVDIPINGLTTPKLTIVGSDKDGKTLV 710
Db 687 DLSSRRSPAGVHFNGIGGKTFNIGNAKAKFLKLP-NAATDPK-----KELPITFNA 739
QY 711 EQVASGNDTKNII-----RGLSPTLPSITNAGVVRTTEQNTITSDKSKAASTG 761
Db 740 NITATGNSDSSVDFDIHANLTSRAGINMDSINITGGL-----DFSITSHNNSNAFEIK 794
QY 762 DIL-----NTGFNLKN-----NSNSVGFVSTYNTVDFIDGNATTAKVTYDETNOTSKVT 810
Db 795 KDLTINATGNSFSLKQTKDSFYNEYSKHAINSHNLTLGCGN-----VTLGGENSSSIT 849
QY 811 YDYNVDEK-TIELTGNGKTNKIGVKTTLTITNANGKATNFSTTDDNALVNAKDIAENL 869
Db 850 GNINITNKANVTLOADTSNSN-TGLKRTLTGNTSVEGNLSLTGANANIVGNLSIAED- 907
QY 870 NTLAKEHTTKGADTALQTFVKYKDGADTDETTITVGKGTONGKTVNPLKL---KGENG 926
Db 908 -----STFKGASDNLAI-----TGFTNNGTANINIKGVVKLGDINNGG 948
QY 927 LVATNKGDTVTFINGINTQSLKAGDSTTLNKGDLSTKNPASNEQIQVGADGVKFAKVDKG 986
Db 949 LNIITNASGKTQIIN-----GNITKEKGLNKNIKAKADAEIQIG-----G 989
QY 987 NSTGTIDGTRITKDOIGFT-----GANGSLDTTKPHL---TKDKLVGVEVLT 1032
Db 990 NISQK-EGNLTISDVKVNTQITIKAGVEGGKSDSEANENLTTQTKELKLAGDLNI- 1047
QY 1033 NTGINAG---GKKITITQSGDITQNSNDA---VTGGRVYDLKTELES---KINSAAKTAQN 1084

Db 1048 -SGFNKAETAKNGSDLTIGNASGKNADAKKVTDFKVKDSKISTDGHNVTLNSEKTSNG 1106
QY 1085 SLHEFSVADEQGN-----HFTVSNPYSDYDTSKTSKSDVITTFAGENGITTKVNGKV 1133
Db 1107 S-----SNAGDNSGCLTISAKDVTVNNVTSHTKINIS-----AAGCNVTTK--EGT 1152
QY 1134 VRVGIDOTKGLTTPKLIVGNNGKGIYJDSKQGN-TITGLSNTLANVNDGAGHAJSQG 1192
Db 1153 T---INATTG--SVEVTAONGTIGKNI-----TSQNVTVTATENL---VTTENAVINATSG 1200
QY 1193 LANDTDKTRAASIGDVLNAGNLOQNGEAVDFSVYDIDGNATTAKTAKTYDDTSKTS 1252
Db 1201 TVNISTKT---GDIKGGIESTSGNVN-----ITASGNTLKVS----- 1234
QY 1253 KVVYDVNDKNTIEVTSDKKLGVKTT-----TLTKTSANGNATKFSAADGDALVLRASDIAT 1308
Db 1235 -----NITGQDVIVTAD--AGALTITTAGSTISATTGNANIT-----TKTGDLNG 1276
QY 1309 HLNTLAGDIO--TAGASOASSASVYVDAGNKKVYIDSTDKKYYQVNDKGQVKNKEVAKD 1367
Db 1277 KVESSSGVTIVATGATLAVGNIS-----GNTVTITA-----DSGKLT----- 1315
QY 1368 KLVAAQOTPDGTLAGMNVKSVINKQOVNDANKKQGINEDNAFIKGLNAAKDTTKNAAV 1427
Db 1316 ---TVGSTINGT-----NSVTTSSQSGD-----IEGTISG---NTVNVTA 1349
QY 1428 TVGDLNVAQOTPLTPAGDTGTTAKKLGETLTAKIKGQDTDNKLTNNIGVAVAGTDFGTYKL 1487
Db 1350 STGD-----LTIGNSAKVEAKNGAATLTAESG-----KLTTGTGSSITSSNGQTTLT 1396
QY 1488 AKDLTNLNSVAGGVRIDEKGISFVDANGQAKANTPVLISANGLDLGGKRISNIGAAVDN 1547
Db 1397 AKDSSIAIGNIAANVTLNTTGT--LTTTGDSKINA---TSGLTINAKDAKLDGGAAGDR 1451
QY 1548 DAVNFKQFNEVAKTVNNLNOSNCASLPFVVVDANGKPIINGTDCGKPKOKAIG--ADGKY 1605
Db 1452 TVVNATN-----ASGSGNVTAKTSSSVNITGDLNTINGLNIISSEGNRTVRLRGKIDVKY 1507
QY 1606 YHANANGVVPKDGKPIITDADKLANLAAGKPLDAGHQVWASLG-----GNSDAIT 1656
Db 1508 IQPGVASVEVEIARVV--LEKVKDLS-----DEERETLAKLGVSAVRFPNNAITVN 1559
QY 1657 LTNIKSTLP 1665
Db 1560 TONEFTTKP 1568

RESULT 12

US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996

QY 1583 NKPINGTGKPKQKAIKG--ADGKYYHANANGVPVDKGPITDADKLANLAHKGKPLDA 1640
 DB 1482 NGLIISSEGRNTRVLRGREIDVKYIQPGVASVEEVIEAKRV--LEKVKDLS-----DE 1533
 QY 1641 GHQVNASLG-----GNSDAITLTNIKSTLP 1665
 DB 1534 ERETLAKGVSAVRFEPPNNAITVNTQNEFTTKP 1567

RESULT 13
 US-08-728-470-10
 ; Sequence 10, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,470
 ; FILING DATE:

CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/302,832
 FILING DATE: 16-MAR-1993
 PRIOR APPLICATION DATA: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-633
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1529 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-728-470-10

Query Match 4.4%; Score 518; DB 2; Length 1529;
 Best local Similarity 21.9%; Pred. No. 9.2e-21;
 Matches 389; Conservative 232; Mismatches 668; Indels 488; Gaps 87;

QY 97 LAKAHANQAIAGGSKPDPNRNOANQKAGSHAKGESIAIGGDVLAEGDASIAIGSDDL 156
 DB 2 MSVVTGATMQVDGKNTIRN-----SYNAIINWKQFN 34
 QY 157 LDNRN-----STNSKYPNGLLSTLTIONTVLRQIRDSNGSKYRRTAEGHASTA 208
 DB 35 IDQEMEQFLOESSNAVFNRVTSQDISQ----LKGILDSNG-QVFLINP-----NG 81

QY 209 MAYAKGHAFANAFG--TRSTAE--GNYSL--AVGLTAKAEKGYTIIAGSNAQAIIYNGALALGAD 265
 DB 82 ITICKDALIINTNGTASTLDISNENIKARNFTLEQTQDKAL-----AEIVNGHGLITVGKD 136
 QY 266 TRVDLDYG-----IALGYSQILNNNNNNNNKAY---VPEGNSNIKSSKA 308
 DB 137 GSVNLIGGKVKNEGVISVNGGSIISLAGOKITTSIDIINPTTYSIAAPENEAINGDIFA 196
 QY 309 TGNGLFSIGSSTIKRK-----IINVAGYEDTDVNVVAQLKAVENIAKLR 352
 DB 197 KG-GNINVRAATIRNKGLSADSVSKDKSGNIVLSAKEGEABEIGGVISAQOQAKG--GKL 254
 QY 353 QITFKGDDNGTGVKKLGETLTI---KGETOQ--ADKLTNNNIGVYTDNNNTGLKVKLAK 407
 DB 255 MIT-----GDKVTLKTCGAVIDLSGEGGETYLGCDERGECKN-----GIQLAK 297
 QY 408 NLSGLETVSTKNLTASEK-----VTVSGNNTAE---LOSGLLTFTPTTNASTDK 454
 DB 298 KTT-LEKGSTINVSGKEKGRRAIVMGDIALDGNINAGGSDIAKTGGFVETSGHDLSIGD 356
 QY 455 TVYGTGDLKFTDNSNTALEDTT---RITKDKIGFSNKAQTVDE-----NKPYPDKDKLK 505
 DB 357 DIVYDAKEMLLDDPDVSIETLTSGRNNITGENQVTTGDCTKESPKNISISKPTLTNTSTLE 416
 QY 506 -----VGNSTLANGLTVNNTGGSNKQIOVGA--DGIK----- 537
 DB 417 QILRRGSYVNIANN--RIYVNSSINLSNGSLTLTKRDGVKINGDITSNENGNITIKAGS 475
 QY 538 PADVNVVNSNAAKFGTTRITEEIGFADADGKVDKPSYLDKQLQGVGVKITKD----- 592
 DB 476 WDVHKNITLTGTGP--LNIVAGDSVAF--EREG--DKARNATDAQITAGTITIVNKDDKQFR 531
 QY 593 -----SGINAGDQKISNVKDATTDDA-----VTYKOLKQVQDADGALQSP 634
 DB 532 ENNVSINGTGKGLKFIANQNNFTKFCDELNISGIVTINQTTKKDKVYNWASKDSYNNVS 591
 QY 635 SIRDEKGOEFFISNLYSNG-----NTPNTPETITEAGENGISISNDIAKGVKVGIDP 687
 DB 592 SLTLNTVQKFTFIKVDGSGNSQDLRRSRRSFAGVHFNIGGKTFNFCANAKALFKLKP 651
 QY 688 INGLTTPKLTVGSKDKGTQLVIEQVASGNDITKNI-----RGLSPTLPSTTNAGG 738
 DB 652 -NAATDPK-----KELPITFNANITATGNSDSSVDFDIHANLTSRAAGINMDSNITGG 704
 QY 739 VRTTEGNTITSDEDKSKAASIGDIL-----NTGFNLKN-----NSNSVGFVSNTYND 787
 DB 705 L-----DFSITSHNRNSNAFEIKDLTINATGNSFSLQTKDSFYNEYSKHAINSSHNT 759
 QY 788 FIDGNATTAKVYDETNOTSKVYDVNVYDEK--TIELTGDNGKTNKIGVKTTLTTLTNANG 846
 DB 760 ILGGN-----VTLGGENSSSITGINITNKANVTLOADTSNSN--TGLKRTLTILGNISV 813
 QY 847 KATNFSTTDNDALVNAKDAENLNTLAKEIHTTKGADTALQTFVKKDKGATDDETTIVG 906
 DB 814 EGNLSLTGANANIVGNLSAED-----STFKGEASDNLNI-----TGFTT 853
 QY 907 KDGTONKTVN-----TLKLKGE-----NGLTVATNKDGTVTFGINTQSLKAGDSTTLNKKD 958
 DB 854 NNGTAN---INIKQGVKLGDDINNKGGNLNITTNAGTOKTILN-----GNITNEKG 902
 QY 959 GLSIRKPASNEQIQVAGDGKVFKAQVDKGNSSSTGIDGTSRITKQDQIGFT-----G 1007
 DB 903 DLNINKIKADAETIQ-----GNISQK--EGNLTISSDKVNITNQITIKAGVEGG 950
 QY 1008 ANGSLDTTPHPL---TKDKLKVGEVEITWTGINAG---GKKTINIOSGDITONSNDAA--V 1059
 DB 951 RSDSSAEANANLTQTKELKLAGDLNI--SGFNKAIEITAKNGSDULTIGNASGNADAKV 1008
 QY 1060 TGGRVVDLKTLES---KINSAAKTAQNSLHFEFSADEQCN-----HFTVSNPY 1105
 DB 1009 TFDKVKDSKISTDGHNVLTNSEVKTSGNS-----SNAGNDNSTGLTISAKDVTNNNV 1061
 QY 1106 SSYDTSKTSVDITTFAGENGITTKVKNKGVVRVIGDQTKGLTTPKLTVGNNGKGIIVDSKD 1165

QY	739	VRTTEQGNITISDEKSKAAISGIDIL-----NTGFNLKN-----NSNSGVFVSTYNTVD	787
Db	705	L-----DFSITSHNRNSAFRIKDLDTNATGNSFLSKQTKDSFYNEYSKHAINSSHNL	759
QY	788	FDGNATTAKVYDNTQTSKVTVDVNVDK-TIELTDGDKTKNGIKGVTTLTTLTTTNRANG	846
Db	760	ILGN-----VTLGSSSSITGNITNKANVTLOADTSNS-TGLKRTLLILGNISV	813
QY	847	KATNFTTNDALVNAKDIAENLTLAKEIHTKTGTDATQATQTKVKKDKGATDDTETIVG	906
Db	814	EGNLSLTGANANIVGNLSIAED-----STFKGEASDNLNI-----TGFTT	853
QY	907	KDGTONGTKVN-----TLKLKE-----NGLVTATNKDGVTVTFTGINTQSLKAGDSTLTKND	958
Db	854	NNGTAN-----INIKQGVKLGQDNNKNGLNITTNASGTFQKTIIN-----GNITNEKG	902
QY	959	GLSIRNPASNEIQVGADGVFAKVDKNGSSTGIDGTSRITKDKQIGFT-----G	1007
Db	903	DLNINIKRADAIEIG-----GNISQK-EGNLTISDDKVNITNQITIRAGVEGG	950
QY	1008	ANGSLDTTKPHI-----TKDKLAVGEVEITNTGINAG-----GKKTINIQSDITQNSNDA--V	1059
Db	951	RSDSSEAEANATIQTKELKLAGDLNI--SGFNKAETAKNGSDLTIGNASGGNADAKKV	1008
QY	1060	TGGRVVDLKTLES--KINSAAKTAQNSLHEFSVADEQGN-----HFTVSNPY	1105
Db	1009	TFDKVADSKISTDGHNVTLNSEVKISNGS-----SNAGNDNSTGLTISAKDVTNNV	1061
QY	1106	SSYDTSKTSDDVITFAGENGITTKVKNKVVRVIGDQTKGLTTPKLVGNNGKGVIDSKD	1165
Db	1062	TSHKTINIS-----AAAGNVTK--EGTT--INATTG-SVEVTAQNGTIKGNI-----T	1105
QY	1166	QON-TITGLSNLNAVNTDNGAGHALSQGLANDTKTRAASIGDVLNAGFNLOGNCEAVDF	1224
Db	1106	SONVTVTATENL--VTTENAVINATSGVNIISTKT-----GDIKGGIESTSGNVN--	1153
QY	1225	VSYDTPVDIDGNATTAKVYTDYDTSKTSKVYDVVNVNKNKTIEVTSDKKLGVKIT--TL	1280
Db	1154	-----ITASGNVLKVS-----NITGQDVTVTAD--AGALTTTAGSTI	1188
QY	1281	TKTSANGNATKFSADGDALVKASDIATHLNTLAGDIO-TAKGASQASSASYVDADGNK	1339
Db	1189	SATGTNANIT-----TKTGDINGKVESSSGSVTLVATGATLAVGNIS-----GNT	1233
QY	1340	VYDSTDKKYQVNDKQGVNDKNEKVAKDLVAQAQTPDGTTLAQMNVKSVINKEQVNDANK	1399
Db	1234	VTITA-----DSGKLTS-----TVGSTINGT-----NSVTTSSQSGD--	1265
QY	1400	KOGINEDNAFIKGLENAKDKTKYKNAAVTVGDLNVAQTPLTFTAGDTGTAKKLGETLTI	1459
Db	1266	-----IEGTISG-----NPNVNTASTGD-----LTIGNSAKVEAKNGAATLTA	1303
QY	1460	KGGQTDNKLTDNNIGVWAGTDGTVKLAKDLTLNLSNVNAGGTTRIDEKGISFVDANGQAK	1519
Db	1304	ESG-----KLTGTGSIITSSNGQTTLTAKDSSIAIGNANAVTLNTTGT--LTTTGDSK	1356
QY	1520	ANTPVLISANGLDLGGKRISNIGAAVDNDVAPNFKOPNEVAKTVNNINNNOSNGASLPFVV	1579
Db	1357	INA--TSGLTITINAKDAKLDAAGSDRTVYVATN-----ASGSGNVTAKTSSSVNITGDL	1409
QY	1580	TDANGRPINTGCKPKQAKTG--ADGKYHYHANANGVPVDKCKPITDADKLANLAHGRP	1637
Db	1410	NTINGLNIISENGNTVRUGKEIDVKYIQPGVASVEEVTEAKRV--LEKVKDLS-----	1462
QY	1638	LDAGHGVVASLG-----GNSDAITLTINIKSTLP	1665
Db	1463	-DEERETFLAKGVSAVRFVEFNNAITVNTQNEFTTKP	1498
RESULT 15			
US-08-038-682-2			
; Sequence 2, Application US/08038682			

Patent No. 5549897	Query Match	4.1%	Score 482;	DB 1;	Length 1536;
GENERAL INFORMATION:	Best Local Similarity	21.0%	Pred. No. 9.2e-19;		
APPLICANT: BARENKAMP, STEPHEN J	Matches	394;	Conservative	257;	Mismatches 649;
APPLICANT: ST. GEME III, JOSEPH W				Indels	572;
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS				Gaps	94;
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS					
NUMBER OF SEQUENCES: 8	QY	1	MNHIVKVFIFNKATGTFMVAEACAKSHSGSSSTAGQVCGSSPVIRLTRVATLAILVICAT	60	
CORRESPONDENCE ADDRESS:	Db	1	MNKIYRLAFESKRLNALVAVSELAR---GCDHST--EKGEKPAEM-KVRLHLAKPLSAM	53	
ADDRESSEE: Shoemaker and Mattare, Ltd	QY	61	LNGSAYAQNNKSIATGTTGNNDNASASNEASTAIGSLAKAHANQAIAIGGSKGDPDRNOAA	120	
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza	Db	54	L-----LSLGVTSIPOSVLAS-----GLOCMQDVVHCTATMQVDGNKTIIRNSV-	96	
CITY: Arlington	QY	121	NOKAGSHAKGKESIAIGGDVLAEGDASIAIGSDLLYLDRN-----STNSKYENGLLS	172	
STATE: Virginia	Db	97	-----DA--IINWKQFNIDQNMVQFLOENNNNSAVFNRTS	130	
COUNTRY: U.S.A.	QY	173	TLIQNHTVLRQIRDSNGSKYRRTAEGHASTAVGAMAYAKCHAFANAG-TRSTAE-GNY	230	
ZIP: 22202-0286	Db	131	NQISQ---LKGILDSNG-QVFLINP-----NGITIGKDAIINTNGFTASTLDSIS	177	
COMPUTER READABLE FORM:	QY	231	SL-AVGLTAKAEKGYTIAIGSNAQAIIYAGALGADTRVDDYG-----IA	275	
COMPUTER: IBM PC compatible	Db	178	NIKARNFTFEOQKDKAL-----AEIVNHGLITVGDGVSNNLIGGKVKNEGVISVNGGSIS	232	
OPERATING SYSTEM: PC-DOS/MS-DOS	QY	276	LGYGSOILNNNNNNKAY---VPEGNGSNIKSSKATGNGLFSIGSSSTIKRK-----	324	
SOFTWARE: PatentIn Release #1.0, Version #1.25	Db	233	LLAGOKITISDIINPTITYSIAAPENAVNLGDIIFAKG-GNINVRATIRNOKGLSADSV	291	
CURRENT APPLICATION DATA:	QY	325	-----IINVGAGYEDTDVAVNAQLKAVENLAKRIFKGDNDGCTGVKKLGETLTI-	375	
APPLICATION NUMBER: US/08/038,682	Db	292	SKDKSGNIVLSAKEGAEGVVISAGNOQAAG-KGLMIT-----GDKVTLTATGAVIDLS	344	
FILING DATE: 16-MAR-1993					
CLASSIFICATION: 424					
ATTORNEY/AGENT INFORMATION:					
NAME: BERKSTRESSER, JERRY W					
REGISTRATION NUMBER: 22,651					
REFERENCE/DOCKET NUMBER: 1038-293					
TELEPHONE: (703) 415-0810					
TELEFAX: (703) 415-0813					
INFORMATION FOR SEQ ID NO: 2:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 1536 amino acids					
TYPE: amino acid					
STRANDEDNESS: single					
TOPOLOGY: linear					
MOLECULE TYPE: protein					
US-08-038-682-2					

Query Match	10.5%	Score 1222.5;	DB 2;	Length 2059;
Best Local Similarity	24.0%;	Pred. No. 3.2e-39;		
Matches 599;	Conservative 368;	Mismatches 855;	Indels 679;	Gaps 115;
QY	39	GSSPVRLTRVATLAILVIGATLNGSAVQAQNSKIAFGTTCGNNDNASASN-EASIAIGSL	97	
Db	13	KEPHLTLFVALSLSLPYTEKNAQVYI--NDGMDGCCQRIVDNGSASGVRTVA----	66	
QY	98	AKAHANOAIAGGSK-PDPRNOAQNOKAGSHAKKESTIAIGDVLAEADSAIAGSDDL	156	
Db	67	TQCEDPWTVTGYARFPGPSGTAAEQGAS-----RNLTGLGSLVYN--SGQGVGVNDV-	117	
QY	157	LDRNSTNSKYPNGLLSTLIQNHVTLRQIRDSNGSQKYRRTAAEGHASTAVCA--MAYAKG	214	
Db	118	-----LNKTYSIRMGSVITMM-TVA-----GTNAIAGSAQSSAADALKASLATAKSG	164	
Ov	215	HFANAGFTRSTAEAGNYSLAVGLITAKAEKG-YTIIAGSNAOAINTYGAALGADTRVDI-DY	272	

Db 165 ARAIAKAGADGVDYVALGSGTAGTGASSIALGILNASAVN-CAYAVGGGALVTPYDQ 223
 QY 273 GIALGYSQILNNNNNNNNNNKAYPEGGNSNIKSSKATGNGL--PSIG---SSTIK-RKII 326
 Db 224 AVALGLNSVA---SFGKGLSGYDPKTKTSTDSAAAMKSTLAAVSIGDVSNTLKRQLS 280
 QY 327 NVGAGYEDTDVAVNAQLKAVENLAKR--QITFKGDDNCT-----GVKKKLGETLT 374
 Db 281 GLAAGTSNTDAVNAQLKVDVDEIASRGNLNTASGANGSNVAPGSSVDLKNPTDKNLITKA 340
 QY 375 IKGGETQ-----ADKLTDNNNIGVVTD---NNTGLKVKLAKMLSGLETYST 417
 Db 341 TGSNDVQENLKNKDKVVTTLAVGDALLNTDGLALGTDVSLSTGLAI-----TDGPAVTA 394
 QY 418 KNLTAASEKVTVGSGNNTAELSGGLTFTPTTNASTDKTVYGTGDLKFTDNTALEDTTR 477
 Db 395 SGIDAGSKV-----ISHVAAGAVSET-----STDAVNGSQLNAVQVQASQP 435
 QY 478 ITKDKLIGPSNKAQTVDENKPYLDKDLKVGNTLNNGLTVNNTIGGSN-KOIQVGADGI 536
 Db 436 VT-----FTGNEGAV-----KRLGQSVTVISGESSTAGTYSGGNLKSVDVDEAAG- 479
 QY 537 KPADVNVVNAAKFQTTTRITEEIECFADADGKVDKSPYLDKQLQVGGVKITKDSGIN 596
 Db 480 ---RIHLQADSPKFGNVINN-----GG-----KISGVT 506
 QY 597 AGDQKISNVKDATDDTDVATYKQLQVQ-----QDADG 629
 Db 507 AG-----TEETDAVNFSQLSISTAVDQGWTLTASGNGSKVASGGTVDLKNTDQ 556
 QY 630 ALQSFISIRDEKGOETTISLNY-SNGNTPTFTTIFAGENGISINDIAKGVKVGIDPI 688
 Db 557 NLTKSGSDNDVFNLSKDFKVDGWTSGT-----TVNNDGKVGSDVALGTTGLTITDQ 612
 QY 689 NGLTTPKLVGSDKDGKLTOLVIEQVAGS---NDTKNITRGLSPTLPSTNACGVFTTQ 745
 Db 613 PAVTASGIDAGSK-----VISHVAAGVYVSETSTDAVNG-----SOLNAVQVQAS-QP 658
 QY 746 NTITSPEDSKAASIGDIINTGFLNKNNSVGFYSTYNTVDIFDGNATTAKVTYDETNO 805
 Db 659 VFTTNEGAVKRS-----LQOSVVISGESSTAGMYSGGNLKSVDVDEAAGRIHLQADSPK 713
 QY 806 TSKVTYDVNVDEKTEIBLTGDNKGNKI-----GVKTT-----TLTTNANGKATNFSTT 854
 Db 714 FGNV--INNGKISGVTAGTETDAVNFSQLKSISTAVDQGWTLTASGNG-----SKV 766
 QY 855 DNDALVNAKDIAENLNTLAKEIHTTKGTADTALQTFKVKKDGATDDETTITVGKDGTONK 914
 Db 767 ASGTVDLKNTDGNLT-----ISKSGSDNDVFNLSKDFKVDGM 805
 QY 915 TVNTLKLKGENGLTVATNKDGTVTFGTINTQSGLKAGDSTTLNKGDLISKNPASNEQIQVG 974
 Db 806 T-----SGTIVVNN-----DGKVGSDVALGTTGLTI-----832
 QY 975 ADGVKPAKVDKNGSSGTIDGTSRITKDKQITGFTGANGSLDTPKHLTKDLKVGVEITNT 1034
 Db 833 -----TDGPA-----VTAS 841
 QY 1035 GINAGGKLTINQSDITONSNDVATGGRVYDLKTELESKINSAAKTAQNSLHEFSVADE 1094
 Db 842 GIDAGSKVISHVAAGVYVSETSTDAVNG-----SOLNAVQVA-----878
 QY 1095 QGNHFTVSNPYSDYTSKTSDDVTFAGENGITTKVKNKGVVYRGIDQTKGLTTPKLVGN 1154
 Db 879 -----SOPVTFGT-----NEGAVKRLGQSVVISGESSTAGTY 911
 QY 1155 NGKCI--VIDSKDGO-NTITGLNTLANVTDGAGHALSGLANDTKTRAA-----SI 1205
 Db 912 SGGNLKSVDVDEAAGRIHLQADSPKFGNVINNGKII--SGVTAGTEETDAVNFSQLSKI 969
 QY 1206 GDVLNAGFNLOGNEAVDFYSTYDITVDF--IDGNATTAKVTYDDTSKTSKVYVDVN---- 1259

Db 970 STAVDOGWTLTASGANGSKVASGGTVDLKNTDGNLTISK-----SGSDNDVFNLSKDFK 1024
 QY 1260 -----VDNKTTEVTSDDKLGKVTKTTLTKTTSANGNATKESAADGDALVKASDIATH 1309
 Db 1025 VDGMTSGTTVVNNDGVKVGSDVALG--TTGL--TIANGPAVTASGIDA-----GSKVISH 1075
 QY 1310 LNTLAGDIOTAKGASOASSASYVDADGNKVIVYDSTDKKYYQVNDKQGVQDNKKEVAKDKL 1369
 Db 1076 V-----AAGAVSETSTDAVNGSQLNAVQVQASOPVTFGTGNE-GAV--KRLSGQSVV 1123
 QY 1370 VAQAQTPDGTGLAQMNVKSVINKE-----QVNDANKKQGINEDN-AFIKGLENAAKDTKT 1422
 Db 1124 ISGESSTAGTYSGCNLKSVDVDEAAGRIHLQADSPKFGNVVINGGKISGVTAGTEETD- 1182
 QY 1423 KNAAVTVGDLNVAQAOTPLTEAGDTGTTAKKLGETLTIKG-----GOTDITNKLTDNNI 1474
 Db 1183 ---AVNFSQLKSI-----STAVDOGWTLTASGANGSKVASGGTVDLKNTDGNL 1227
 QY 1475 GVAGAGTDGFTV--KLAKDLTNLNSVNAGGTRIDEKGISF-----VDANGQAKANTPVLSA 1527
 Db 1228 TISKSGSDNDVFNLSKDF-KVDGMTSGTTVVNNDGVKVGSDVALGTTGLTIANGPAVTA 1286
 QY 1528 NGLDLGCKRTISNIGA-AVDD--NDAVNFKOFNEVAKTVNLLNNOSSNGASLPFVVTDANG 1584
 Db 1287 SGIDAGSKVISHVAAGAVSETSTDAVNGSQLNAV-----QVQASOPVTF-- 1331
 QY 1585 KPINGTDGKPKQKAI-----KGADGKYIHANANGVPVDKDGK---PITDADKLANLA 1632
 Db 1332 ---GNEGAVKRLSGQSVVISGESSTAGTYSGGNLKSVDVDEAAGTIHLQADSPKFGNV 1387
 QY 1633 AHGPLDAGHOVVASLGG--NSDAITLTNLIKSTLPQIDTPTNGNAGAOQSLPSLSAAQ 1690
 Db 1388 ---INNGKISGVTAGTEETDAVNFSQLKSISTAVD-----QGWTL-TASGAN 1431
 QY 1691 QSNAA--VKDVLNVGFNLQTNHNVQVDFKAYDITVNFVNGTGADITSVRSADGTMNITV 1748
 Db 1432 GSKVASGGTVDLKNTDGNLTISK-----SGSDNDVFNLSKDFKESITV 1476
 QY 1749 -NTALAAATDDGDNVLIKAKDGKFKYKADLMPNGSLKAGKSASDAKT-PTGLSLVNPNAKG 1806
 Db 1477 GNTQL---DKDG---VKYSSNVLLDSNELV---ITSHSSTSSVKTLANGESVNVRTVVN 1526
 QY 1807 G---STGDAVALNLSKAVFKSKDGTITTTVSSDGISITQGDKNSSITLSKDLNAGVKVI 1863
 Db 1527 GDGVNIDVVVNDL-----GLSIVG--GASLTLS--GINAGSHKI 1563
 QY 1864 SNVKGKTDTDAAVVOQLNEVRNLLGLG-NAGNDNADGNQV---NIADIKK-DPN---SG 1915
 Db 1564 TNVTAGTEDTDVAVNFSQLKSVDVSEAVDKGWTLTASGANGSKVYSGGTVDLKNTDGNLAI 1623
 QY 1916 SSSNRTVI-----KAGTVLGGKGNNDTEKLTATGGVGVVDKDGNGANG---DLSNVVW 1964
 Db 1624 SGSDNDVFNLSKDFKVDVETAG-----NTVVNTDGVKVGSDVSLGAMGLFIANGPSVTA 1678
 QY 1965 KTKQDGSKKALLIATYNAAGQTNVLTNPA---EATDRINEQGIREFHVNDGNQEPVQGR 2021
 Db 1679 SGFNAGDK---VISHVAVGMADTDVAVNSQLQAQVQSVTVKATRYISTNDG---TOGG 1731
 QY 2022 NGIDSASGKHVAIGFQNAKADGEAAVIGRQTA-GNOSTAIGDNGAQTGOSTAIGTG 2080
 Db 1732 NYDGCATGSKAIAAGVGTQASGEGAAVVGSAAGSGKSGTAIGHNAIASADGVALGDG 1791
 QY 2081 ---NVVTGKHSG-----AIGDPSTVKADNSYSVGNNNQIFDAP-----2115
 Db 1792 AKDGGGAESYTGKISGVQNNVTGVISVGDAAKGETRSISNVADAKEAMDVNLQOLDV 1851
 QY 2116 ---QT-----DVEFGV--GNNITVTES---NSVALGNSAISAGTHACTQAK 2153
 Db 1852 AOKSNLQTTDDMRHEINNIENIEDVFKITGDSASSVKMGVGNMAIGTNAVVS-GTESVALG 1910
 QY 2154 KSDGTAGTTTTAGATGTVKGFAGQTAAGVAVSVGASGAERRIQNVAAAGEVSATSTDAVNGS 2213
 Db 1911 NTNVSADNNAVIG-NGSVADRA-----NSVSVSGSGSERQVTVNVAAG---TADTDVAVNS 1961

[illegible]


```
Qy 1556 NEVAKTVNNLNQNSGASLFFVVTVDANGKPIGTGDKPQAKGADGCKYHYHANANGVPV 1615
Db 1171 ---TVPOLNATCAVATTP-----TKYFHANS-----1194
Qy 1616 DKDKPITDADKLA-----NLAHKGKPLDAGHQVVA-SLGGNSDAITLTNTKSTLPQIDTP 1670
Db 1195 -----TEEDSLAVGTDSLAKGAKTIVGDKGIGIGYGAVVDANALNGIA-----1238
Qy 1671 NTGNANAGQAQSLSLSAQNSNAASVKDVLNVGNFQTNHNVDFVKAYDTVNFVNGTG 1730
Db 1239 -----IGSNAQVVIH-----VNSIAIGNS-1257
Qy 1731 ADITSVRSADGTMSTNITVNTALAAATDDGNVLIRAKDKFKYKADDLMPNGSLKAGKSASD 1790
Db 1258 ---TITRCAQ-----1264
Qy 1791 AKTPTGLSLVNPNGAGKSTGDAVALNLSKAVFKSKDGTFTTTSVSSDGISTQGDNSISIT 1850
Db 1265 ---TNVTAYNMDAPQNSVGE-----FSVGSAD-----1288
Qy 1851 LSKDGLNVGGKVIENVGKGTDKTDAANVQQLNEVRNLILGLGNAGDNADGQVNIADIKK 1910
Db 1289 -----GQRIITNVAAGSADTDVAVNGQL-----KVTDQV 1318
Qy 1911 DPNSSSSNRRTVIRAGTVLGGKGNNDTEKLTATGGVQVGVKDXGNAGDLSNVWVKTKQDKG 1970
Db 1319 SONTOSITN-----LDNRVTNLDNRVT-NTENGI-----GDIV-----1350
Qy 1971 SKKALLATYNAGQNTYLTNPAEADRINEQGRIFRHHVNDGNEPVPVQGRNGIDSSASG 2030
Db 1351 -----TGSTKPKYKTN-----TDGVDASAQ 1371
Qy 2031 KHSVAIGFOARADCEAAVAICRQTOAGNOSTAIGNAQATGDSIAIGTGNVTVGKHSGA 2090
Db 1372 KDSVAI-----GSGSIAADN-----SVALGTGSVAT-----1398
Qy 2091 IGDPTVKADNSYSGVNNNQPIDATQDFVFGNNITVTESNVALGNSAISAGTHAGT 2150
Db 1399 -----EENTISVGSST-----NORRIT-NNVAAGKNA-----TDAVN 1428
Qy 2151 QAK-KSDCTAGTTTAGTGTGKGFAGTAVGAVSVGASGAERRTIONVAAGEVSATSDA 2209
Db 1429 VAQLKSSBAGVRDTRKADGSID--YSNITLG-----GGNGGTRILSNVSA6---VNNNDV 1479
Qy 2210 VNGSOLYKATQCIANATN-----ELDHRIHQENKANAGISSAMAMASMPQAYIFCRSMVT 2265
Db 1480 VNYAQLKQSVQETKQYTDORMVMDNKLKSKTESKLSGGIASAMAMTGLPQAYTPGASMAS 1539
Qy 2266 GGIATHNQGGAVAVGLSKLSDNGQWFKINGSADTQGHVGAAGAGFH 2314
Db 1540 TGGGTNGESAVAGLVSMVSGRWRVYKLOGSTNSQGEYSAAALGAGTQW 1588

RESULT 3
A82615
surface protein xfl981 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82615
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82615
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1190 <SIM>
A:Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
```

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
J.D.; Junqueira, M.L.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
Chado, M.A.; Madeira, A.M.B.N.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
A:Authors: Martins, E.M.F.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
F.G.; Nunes, L.R.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
Rodrigues, V.; Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Vojtko, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XFL981

Query Match 6.48; Score 743.5; DB 2; Length 1190;
Best Local Similarity 24.1%; Pred. No. 3.3e-21;
Matches 352; Conservative 188; Mismatches 481; Indels 441; Gaps 67;
Qy 1010 GSLDTTKPHLTDKDLKYGE-----VEITNTGINAGCKKITNTQSGDITQNSNDVATGGR-- 1063
Db 13 GGIEGVDP-LFLPKYKICQSLQHAVMTSA---ASSKK-----GTQPRSRNMAKRS 62
Qy 1064 -VYD-----LKTELESKINSAAKTA-----ONSLHEFSVADEQGNHFTVSNPY 1105
Db 63 KLHRRQLHIVLLTFLAASGTGKVAQVYVNSDSTENCVEILGDSQTSQTSFIHSASNDK 122
Qy 1106 SSYDTSKTSDDVITFAGEN---GITTKVKNKGVVRVIGDQTKGLT-TPKL-TVGNNGK--- 1157
Db 123 CKPFTQTEYSLFDYRNLVGLGSLYVNEG--KLGLVDISGATSMRLGSLTATNGWSAGI 180
Qy 1158 -GIVIDSKDGQNTITGLSNTLANVTNDG--AGHALSQGLANDTKTRAASIGCDVLNAGFN 1214
Db 181 DSIAGSGQSGKT-----DQNTSGATVAQGL-----RSIAIGTTARS--- 217
Qy 1215 LQNGEAVDFVSTVDFVDFIDGNATTAKVTYDDTSKTSKVYDVNDKNTTEVTSDDKLG 1274
Db 218 -----QSQDASIGTCASITGTFNFAI----- 237
Qy 1275 VKTTTLTKTSANGNATKPSAADGALVKASDIATHLNTLAGDIQDTAKGASQASASVYD 1334
Db 238 -----AIGNGALTSIANGALGASSVTIR-----GGVALGQGSALATASGI--- 279
Qy 1335 ADGNKVIYDSTDKYKYQVNDKGQVKNKEVAKDLVAQAQTPDGTFLAQMNVKSVINKQV 1394
Db 280 ---TGYPDPVTKS-----TSTLSTSMRSTLGAVSI 306
Qy 1395 NDANKQGINEDNAFIKGLNAAKDTKTNAAVTVGDLNVAQOT-----PLTFAGDTGTT 1449
Db 307 CNITVSS---TSQTRQLTGCLAAGRSQTD---AVNVAQLKLLAESVGGGWNLTASGANSN 359
Qy 1450 AKKLGETLTIRGGQTDNKLTDNNINVGAVAGTDG--FTVKLAKDLTNLNSVNAAGGTIRDEK 1507
Db 360 V-ALGESVDLKN-----SDGNLLITKTTDSNDVTENLATAL-KVDSLATGTNTAMTDD 409
Qy 1508 GISF-----VDANGQAKANTPVLSANGLDLGGCKRISNLGAADDNDVNFQNFVAKTV 1562
Db 410 GVTVGKRVTLDSGTGLVTAEGPSVSISSGINAAGQKTMNVTGTADTDVNFQOLQAVSDTA 469
Qy 1563 NNLNNSQNSGASLPEVVTGDKPKINGTGPKQAKGDKGYIHANANGVPVVDKDKPI 1622
Db 470 SKGWNLLASGYN-----SSNVAPGASVD-----LKWTDGNLLITKAIGI----- 508
Qy 1623 TDADKLANLAAGKPLDAGHQVAVSLGNSDAITLTNTKSTLPQIDTPTNTGNA-NAGAAQ 1681
Db 509 ---NDVTENLAT---ALEADSLTGTNTAMTDTGTVGS-----NVTLGSTGLVITDGPVS 557
Qy 1682 SLPSLSAAQ---SNAASVKDVLNVGNFNLQTNHNOVDFVKAYDTVNF-----VNGTGADI 1733
Db 558 TSSGISAGNQNKITNVAAGTADT-----DAVNFQOLQAVSSTASKG 597
Qy 1734 TSVRSADGTMSTNITVNTALAAATDDGNVLI-----KAKDGKFKYKADDLMPNGSLKAGKSAS 1789


```
Db 598 WNLASGANSNVVPGESVDLKNDSGNLLITKTDSNDVTFNLATALKVD-SLTTGNT-- 654
QY 1790 DAKPTGLSLVNPAGKSGTDAVALNLSKAVFKSKDGTITTTVSSDGSIOGKDNSSI 1849
Db 655 -AMTIDGVT-VGSNVTLGSGT-----LVITDGPVS----- 682
QY 1850 TLSKDLGNVCGKVISNVGKTKTDAANVOQLNEVR-----NLGLG-NAGN----- 1895
Db 683 --TSSGISAGNOKITNVAAGTADTDAVNFSQLQAVSSSTASKGNLLASGANSNVAPGES 740
QY 1896 ---DNADGNVNIADIKKDNSSG-----SSNRTVIRAGTVLGG 1931
Db 741 VDLKNTDGNIV---TSKESGNDVLPNLSSSLKDLKLTVDGTVMTNGVTGSGVTLGS 796
QY 1932 KG--NNDTEKATGGVQGVGDK-----DGNANGDLSNV-VWTKQDKGSKKALLATYNAAG 1983
Db 797 MGLVITDGPVSTSSGINAGSQKITNVAAGTADTDAVNLSQLNTAMAGS----- 844
QY 1984 QTNVLTNNPAEATDRNEQIRFEHVNDGNQEPVVOGRNGIDSSAGKHSVAIGFOAKAD 2043
Db 845 -----GAKSVHYISTYDGG-----TOGGNYNGDGTGTRSIAGVGTLAS 884
QY 2044 GEAATIGROTQA-GNOSIAIGDNAOATGQSTAIATG-----NVVTGKHSG----- 2089
Db 885 AEGATVGSAAASGKSTAGRNAVASADGSVALGDGAKDGARGAESYTGKYSGLQNTT 944
QY 2090 -----AIGDPSTVRKADNSYSYGNNNQFIDAT---QTDVFG-----VGNNTI-----TVT 2129
Db 945 VGTVSGDASKGETRIVSNVADAKEATDAVNLQDLQDRAVDQADANRYVDNKTESLSEGTFFV 1004
QY 2130 ESNVALGNSA--ISAGTHA-----GTOAKKSDGTA-GTTTATAGATGVKGFAGOTAVG 2181
Db 1005 KVNSL---NNSATPIAAGVDATAIGVGATASGADSIAMGNKASASADNAV-----AIG 1054
QY 2182 -----AVSVCASCAERRIONVAAGEVSATSDAVNGSOLYKATOGTANATNELDHR 2232
Db 1055 NHSVADRANTVSGSAGSERQVNTNVAAG---TADTDVAVNSQL---NQGLITAKQYTDGV 1108
QY 2233 IHONENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGGAVAVGLSKLSDNGQWVF 2292
Db 1109 VGSILRRDTDGVAATAATANLPOAYIPGRGMTSVGVSSYRGQSAIAGVSSVSESGRWF 1168
QY 2293 KINGSADTQGHVGAAGVAGPHF 2314
Db 1169 KFGSANTRSQVGTGAGVGYQW 1190

RESULT 4
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; MID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20615

Query Match 5.7%; Score 668.5; DB 2; Length 5188;
Best Local Similarity 21.6%; Pred. No. 1.7e-17;
Matches 606; Conservative 290; Mismatches 1017; Indels 895; Gaps 141;
```

```
QY 29 GSSSSTAGQVGSPPVIRLTVATLAILVIG-ATLNGSAYAQNNNSKIAFGTTGNNDNASAS 87
Db 283 GGYTYTATVAGDGSWTTTTPAADLSVLRDGDTAQVASV-----STINGNTASAT 331
QY 88 NEASI-----AIGSLA-----KAHANQAIAGGSKPDPNQAANQKAGSHAKGRESI 134
Db 332 HAYSDVATATPALTINTIATDDILNAAEAGNPLTISGSSTAEAGOTVTVTL-----NGV 384
QY 135 AIGGDVLAEGDASIAIGSDDLYLDRNSTNSKPNGLLSTLIQNHTVLRLQRDSNGSOKYR 194
Db 385 TYSGSVQADGSWSVSLPTADL---SNLTASQY-----TVSASVSDKAG----- 424
QY 195 RTAAEGHASTAVGAMAYAKHGFANAFGTRSTAEGNYSLAYLTAKEKGYTIIAGSNAQA 254
Db 425 -----NPASANHGLAVDLTVPV---LTINTVSGDDI 452
QY 255 INYGALGADTRVDLDYGIALGVGSOILNNNNN-----NNKAYV-----PEGN 299
Db 453 INAA-----EHGOAL-----VSGSYGGGAGDVITVTLASKYTYTMLDASGN 495
QY 300 GS-NIKSSKATGNLFSIGSSTIKRKIINYGAGYED---TDAVNVAQ----- 342
Db 496 WSVGVPAADVTALG---SGPQTITAAITDAAGNSDDASRTVTYNLAAPTIGINTIATDDV 552
QY 343 LKAVERNIAKQITFKGDDNGTGVKKLGETLITKGGETQADKLDNNNIGVVTNNNTGLK 402
Db 553 IKATEKGADLQIT-----GTS-NOPAGTITV-----TLNGQNYTATTDSNGXWS 596
QY 403 VKLAKNLGLETYSTKNTLASEKVTGSGSN-NTAE---LOSGLTFTPTTNASTDKTVYG 458
Db 597 ATVP--ASAVSALGEANTVTANVTDTAGNSNSASHNVLSALPAVTINAVATDDLIINA 654
QY 459 TDGLKFTDNSNT-----ALEDTTRITKDKIGFSKNAGTVDENKPYLDKDKLVGNST 510
Db 655 AE---SGNAQITSGQVYTGAAQDVTVT---LGGNYITATVQSNLSW-SVDVPAADIQA 706
QY 511 LNNGLTVNNTI-----GGSNKQIOVGA-----DGKIFADV----- 541
Db 707 LGNGDLTVNASVTVNGVNTGSGSRDITIDANLPGLRVDTVAGDDVINSIEHQALVITGS 766
QY 542 -----NVAVSNAAKFTGTRITEEIECFADADCKVKKSPYLDKLOLVGVKITK 591
Db 767 SSGLTAGTALTVEINNTV-YGATVTL-----ADGTSVSGVPAVDVSNMPAGTVNITV 816
QY 592 DSGINAGDQKISNVKDATDDTD--AVTYKQKOVQDADGALQSFSIRDEKQGQFTISNL 649
Db 817 -SGTNSAGTTSTITHPVTVDLAGVAITNL-----SGDDVINAV-----EKGETLVVSGS 866
QY 650 YSNGNTPTNPTETITFAGENGISINDIAKGVKVIDPINGLTPTPKLTVGSDKDKGTQOLV 709
Db 867 TSGVEAGQTV-TVTFGGKNYTTTVE--ANGSWTVNVPPADLAALP-----DGAGNVQAS 917
QY 710 IEQVASGNDTKNIIRGLSPILPTLTNAGVYRTTEGQNTTISDE-----DKSK 756
Db 918 VSNINGNSAADRAYSVDATAPLVTI-----NTIASDDILNVSEACAGITISGTT 967
QY 757 AASIGDILNTGFLKNKNSVSGFVSTYNTVDFIDGN-----ATTAKV 798
Db 968 TAQAQGLTV--TLNNN-----TYQTTVLADGTVSNVPPADLSGLTASSYTVTATV 1017
QY 799 TYDETNTQTS---KVTYDVNVDEKTIETLTGDKGNTKI-----GVKTTTITTT 842
Db 1018 SDRAGNPASADHALVVDITAPDLTINTVAGDDIINAJIEHQALVVSSTGTSGAAAGDVTV 1077
QY 843 NANGKATNFSTDDNALVN-----AKDI---AENLNTLAKEIHTTKGTADTALQTFVK 893
Db 1078 TLNCK--NYTTT-LDASGNWSVGIIPAADVATALATGTSOTITASLSDRAGNSDSTTHDVTVD 1134
QY 894 KDGAT-----DD-----ETITVKGDTGKGTNGKTV----- 916
Db 1135 LSGPTLTINTVSGDDIINAAEIVVAQITISGOVGTAVAGNTVITVIGNQYNATVQSDLS 1194
QY 917 -----NTLKLKGENGLTV-----ATNKDGTVTFTGINTQSGLKGAGDSTTLNKGDLGIK 963
```


C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: B71704
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, T.;
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: B71704
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2340 <AND>
A;Cross-references: GB:AJ235271; GB:AJ235269; NID:G3868717; PIDN:CAAL14908.1; PID:G386100
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: sca3; RP451

Query Match 4.9%; Score 571.5; DB 2; Length 2340;
Best Local Similarity 21.7%; Pred. No. 3.1e-14;
Matches 554; Conservative 295; Mismatches 935; Indels 773; Caps 133;

Qy	59	ATLNGSAYAOONNKIAFGT--TGNNDNASASNEASIAIGSLAKAHANQIAIGSGKPDPPN	117
Db	14	ASLCGTLFTNSN--ATGTIIPNNGSVSLNTDAGLVGVFNNGDITQ--TVNGGREIKISA	69
Qy	118	QAANOKAGSHAKGESIAIGGDVLAEGDASIAIG---SDDLVLDRNSTNSKYPNCLLST	173
Db	70	DKANAIGGTLTKELPDREG---VEVSQNVSIPLUNAGEDL-----NTNF--GPLKF	117
Qy	174	LIONHTVLROIRDSNGSQRYRTAAEGCHASTAVGAMAYAKGHAFANAGFTRSTAEGNYSIA	233
Db	118	ISNNVTSI-----ITGVGKTFESNIDFA--GKNATLQINKOL--	152
Qy	234	VGLTAKAEKQYTTAIGSNAQAQINYGALALGADTRVDLDYGIA---LGYSGQILN--NNNN	289
Db	153	-NITTTID-----NTVAGNNGSITFEGS-----GIISNHIGYTNLSLIGINVGNG	195
Qy	290	NKAYVPEGNGSNIKSKATGNGLFSIGSSTIKRKIINVGAGVEDTDVNVQAQLKAVENL	349
Db	196	EAKYIAPAENNTIINAKN-----INLTHNNSILTL	225
Qy	350	AKROI--TFRGDDNGT-----GYKKLGETLTIKGGETQADKL--TDNNNIGVVYTDNN	398
Db	226	CDGNITTLKGINWTTIDGQGLNLAYDLGSSSIITGDIGNIGSLDTINVLGSGATFNS	285
Qy	399	TGLKVKLAKNLSGLETVSKNLTASEKVTVGS--GNNTAELOS-----	439
Db	286	TILK---ATNINLKHNTSLNLD--DNIIYVIGNIKNNKDIILFKVGHGTNLDNEMIIPAP	341
Qy	440	---GGTLFTPTTNASTDVTYGTCLKFTDONSNTALEDTTRTKDKIGFSNK---AGTV	492
Db	342	QKTHGYLNF--KGNATLNGNINLNLTKESGGHGKTLNQGNTKVDNLVFAOSVLDSGTI	399
Qy	493	DENKPYLDKDKLVGNSSTLNNGLTVN--NTI-----GGSNKQIQVGADGKIFADVNVNV	545
Db	400	SVN--GLLDTDCVTFNNSVNGGPTLIINAKNTISAKLLNATKAKIQIN-----ANLTMMH	452
Qy	546	SNAKEGTTRITEEETGFADA--DGKVDKKSPLYLDKKQLQGVGVKTKDSG-----INAG	598
Db	453	PSAGDISDIRIADNTIYTDIAKNGVN-----LLNNNAKIIFEGADSMALALINTG	502
Qy	599	DQKISNVKDATDTDVATYKQLKQVOODADGALQPSIRDERGQEFITISNLYSNGNTPTNT	658
Db	503	-----VTADRTFTIYNLNQSGNDEGVKIEAIK-----	532
Qy	659	FEITITAGENG--ISISND-----IAGKVKVGDIPDINGLTTPKLTVGSDKDKGTQLV	709
Db	533	-KVITIANOSGPYTIQODNTHRLKELIVECAGDIID--DTITFKLLSINSTSGTITFNT	589
Qy	710	IEQVASGNDTKN-----IIRGLSPTLPSTINAGGVYRTTEQGN---TITSDEKSKAAASIG	761
Db	590	LDLGAGGNTAFGKHGLVNVNGVTGSIITSENQOGLITINSNGITGVIGTINLGLKLVNIG	649
Qy	762	-DILNTGFNLKNNNSNGVFSTYNTVDFLDGNNAKTAVTYDETNQTSKYTYDVNVYDEKTI	820

Db	650	ADPVTCSANY	---FASVALTNFSSVILILADGVTLTGCVT---THNNTKGVLSLG-----	697
Qy	821	ELTGDNCKTKIGVKVTKITLTTTNNANGKATNFS-----TTDNALVNAK	863	
Db	698	--TGSN--ITQIGTGNAALEKINIGAGASNIDSNIIYAGSVLTDTOTSELTLNDVVVNSN	754	
Qy	864	DIAENLNTLAKETHHTTKGRADTALQFFVKVKOGATDDEFTIVGKDGTON--OKTWTNKLK	922	
Db	755	IITTAGNNSKLIFTNGGGI-----TGNIGANGAALQEVWF---NGTNNIGGTANS---	802	
Qy	923	GENCLTVAITNKKDCTVTFGINTOSGLKAGDOSTTLNKDG-----LSIKNPASNEQIOVGADG	977	
Db	803	--QNFVVAHSAANVJTGLTT--GALKYKDTGTLIIAHGLVGDIDFNNAKAG-----	849	
Qy	978	VKFAKVDKGNSSGTIGDTSRIITKDQIGFTGANGSLDTPKPHLTDKDKLKGVEIT--NTGI	1036	
Db	850	-KFLTGD---GAMIDGSLCN-----GGVACTLDF-----IGDGNVTQONIGA	887	
Qy	1037	-NAGGKKITNIQSDITON---SND-----VTGGRVYDLKTELESINSAAKTAQNSL	1086	
Db	888	DNANSISTINIO--GDNTKNVTIANDIFYDNIHHTNGGILQL-----	927	
Qy	1087	HEFSVADEQGNHFTVSNPYSSDYTSKTSDVITFAFAGENGITTKVKNKGVVRVGIDOTKGLTT	1146	
Db	928	-----GGNLTHN-----IDF--GANGSTLEFN-----	948	
Qy	1147	PKLTVGN--NGKGIVIDSKDGQNTITGLSNTLANVYNDGAGHALSQGLANDTKT--RAA	1203	
Db	949	-----GNNTYNLNAIIV---NGQN---GTLNFTNLK-----ASDQITCTVKII	986	
Qy	1204	SIGDV--LNAGFNLOGNGEAVDPVSTVD--TVDFIDGNA--TTAKVTYDPTSKTSK-----	1253	
Db	987	NIGQIGTPQNFITQVNNKNTLTVSSVNSINFGDANSQILIASPVDQTIKFINNLNETGG	1046	
Qy	1254	--VYDVNVNDKNTIEVTSKDLGVKTTTLTKTSANG-----NATKFSAA	1295	
Db	1047	GIITLDSNGNLLTISGNGGIKLSGKNELLSNIIKGVKVTVTNDLDOIHQHLNINNGALF	1106	
Qy	1296	DGALYKASDIATHLNTLAG-----DIOTA-----KGASQAS--	1327	
Db	1107	DDQSLTSAKTKINICTGVAGGATYTLDAINDNFDLNTSGMFKHODSILELKNSSNTNDH	1166	
Qy	1328	--SSASVVDADGNK--VIXYDSTDKKYQVNDKQGVDDKNKEAKD--KLVAQAQTPDGTLA	1381	
Db	1167	TIITLSALDPGNQFGIILITDNTKLTIDNNGNVAYTLGTANHMLKQLTFASIDNGAIA	1256	
Qy	1382	QMVVKSVINKEQVNDANKKQGINEDNAFTKLENAAKDTKTNAAYT--VGDLNAVAQTPPL	1440	
Db	1227	--LKGVINVENTYTLNKOIELENNVANY--LFN-----KNTTYTATGNGH---HV	1270	
Qy	1441	TFAGDTGTTAKKLGETLTIKGGTOTDNK---LTDNNIGVVGADTGFVKVL---AKDLTN	1493	
Db	1271	DFQGNAGVI--NLNDIDIEIDGSVTSGVNGVTLNFGSKVTGLINNIVMLQAGADVSL	1328	
Qy	1494	LNSVNAAGTTRIDEKGISFVDANGQAKANTPVLUSANGLDI-----GKKRISNIG--RAVD	1545	
Db	1329	SASGNTSYITEIQNGNNNNTTFAANSHLTDTINKTGQDLNLVFTNGGVSIGSIGANAAVG	1388	
Qy	1546	D-----NDAVNFKQNEVAKTVNNLNQNSGASLP---FVVTDIANGKPIGTGDKPKQA	1597	
Db	1389	DIILINAGSVNF-----SNTLKSGLNIVISDGATMQVNNVTATDISCKNAN-----	1433	
Qy	1598	IKGADKYYHIANAVPVVDKCKPTDDAKLANLAHCKPLDAGHQVVVASLGGNSDAITL	1657	
Db	1434	-----NGPLKLNHNITITSTLGNNAIG-----TIEVANNDVITITG	1471	
Qy	1658	TNIKSTLPQIDTPNTCNANAGAAQSLPSLSAAQOSNAASVKDVLNVPGLQTNHQNQDFV	1717	
Db	1472	T-----LQAGNIHFNSATQATL--TLGASQ-----VTNITTAGNNIITLE-----V	1512	
Qy	1718	KAYDTVN--FVNGTGDADITSVR--SADGTMNSNITVN-----TALAATDDGCVNLIKADG	1768	

Db 1372 SVTATNPT-----GNSAT--ATQTIV-----VDTTAPT 1397
Qy 1019 LTKDLKVE-----VEITNGINAGKKIINIOGSDITQNSNDVATGGRYVD 1066
Db 1398 VTINAIAVDDIINAVEAGSPVAVSGTTGVEDG-----QVVTVTIDGNTYATVGTNAWT 1452
Qy 1067 LKTELESKINSAAKTAQNSLHEFSVADGQNHFTVSNPYSSYDTSKTSVDITFACENGIT 1126
Db 1453 FNIPIADIANFEATEEVA-----TVSDLAGNPATPATRNIITVDTTAPTVTINAIAVDDII 1508
Qy 1127 TKVNGGVVRVIGDQTKGLTTPKLTGNNNGK--GIVIDSKDQONTITGLSNT-----LANV 1180
Db 1509 NAVEAG-SPVAVSGT-----TTGVEDGQVVTVIDGNTYATVGTNAWTNFIPIADI 1559
Qy 1181 TN-----DGAGHALSQGLANDTKTRAASIG-DVLNAGFNLOGNEAVDFVST 1227
Db 1560 ANFEATEEVAIVTSVSLAGNPATPATRNIITVDTVAPAVTIDSISDDTGAQAN-----DFITN 1615
Qy 1228 YTVDFIDGNA--TTAKVTYDDT--SKTSKVVDVND-NKTIETVSDKKLGKVTITL 1280
Db 1616 DDTLVF-NGTAEADSTVVVSLDGIIGVTANGAGETWLDYTGTLADGDVYELSVTATNP 1674
Qy 1281 TKTSANGNATKFSAADGDALYKASDIATHLNTLAGD--IQAKGASQASSASYVDADGN 1338
Db 1675 TGNSTATATQ-----IVDVTAPTVTINAIAVDDIINAVEAGSPVAVSGTTGVEDG 1726
Qy 1339 KVIYDSTKKYQVNDKQVNDKNEKVAQKLVQAQTPDGLTLOMNVKSVINKEQVNDAN 1398
Db 1727 QVVTVIDGNTYATVGT-----NAWTFNIPVADIANFEATEEVV 1766
Qy 1399 KKGINEDNAFIKLENAKDTKTNA--AVTVGD-LNAV-AOTPLTFAGDTGTAKKL 1453
Db 1767 ATVSLAGNPATPATRNIITVDTTAPTVTINAIAVDDIINAVEAGSPVAVSGT--TTGVED 1824
Qy 1454 GETLIK-GGQTDNKLTDN-----NIGV-----VAG-----TDGFT 1484
Db 1825 GOVVTVIDGNTYATVGTNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATRNI 1884
Qy 1485 VKLAKDLTNLSVNAAGTRIDEKISFVD-----ANGQAKA-NTPVLSANGLDLGG 1534
Db 1885 VDTVAPVTTIDSIS-----BDTGAQANDFITNDTDLVFNGETAEDSTVVVSLDGIETG 1938
Qy 1535 KRISNIGAAVDDNDVAFKQNEVAKTVNNLNQNSCASLPFVVVTDANGRP----- 1586
Db 1939 VTANGAGETWLDYTGTLAD-GDYELSVTATNPNGNSATATQTIIVDVTAPTVTINAIAV 1997
Qy 1587 ---INGDGKPKQAKGA-----DGKYYHANAG-----VPV----- 1615
Db 1998 DDIINAVEAGSPVAVSGTTGVEDGQVVTVIDGNTYATVGTNAWTFNIPVADIANFEA 2057
Qy 1616 -----DKDGRPITDA-----DKLANLAHGKP----- 1637
Db 2058 TEVVATVSDLAGNPATPATRNIITVDTTAPTVTINAIAVDDIINAVEAGSPVAVSGTTG 2117
Qy 1638 LDAGHQVVASLGGNSDALTLLNIKST--LPQIDTPN-----TGNMAGQAO 1681
Db 2118 VEDGQVVTVIDGNTYATVGTNAWTFNIPVADIANFEATEEVVATVSDLAGNPATPATR 2177
Qy 1682 SLPSLSAAQQAASVSKDVLNNGFNLOTHNOVDVFKAYDTVFNFGTG-ADITSVRSAD 1740
Db 2178 NITVDTVAPAVTIDSIS-----DTGAQANDFITNDTDLVF-NGTAEADSTVVVSLD 2228
Qy 1741 G-TWSNITVNTALATDDGNVLKADGKFKKADDLMPNSGLKAGKASDAKTPGLSL 1799
Db 2229 GIEGTVTANGAGETWLDYTGTL--ADGDY-----ELSVTATNPNGNSATATQ-----I 2277
Qy 1800 VNPNAKGSTGDVALNMLSKAFVSKK-----DGTITTTTVSSDGISIQCKDNSSITLSKDG 1855
Db 2278 VVDITAPTVTINAIAVDDIINAVEAGSPVAVSGTT-----GV-----EDGQVVTVIDG 2327
Qy 1856 LNVGKVTISVNGK-GTKDTDANVOQLNEVRNLLGLGNAGNDNADGNQVNTADIKKDPNS 1914
Db 2328 NTYTATVGTNAWTFNIPVADIANFEATEEV-----VATVSDLAGNPAT 2370

Qy 1915 GSSSNRTV-----IKACTVLGGKGNNDTEKLATCGGVQVG-----VD 1950
Db 2371 PATRNIITVDTTAPTVTINAIAVDDIINAVEAGSPVAVSG-----TTTGVEDGQVVTVT 2423
Qy 1951 KQGN-----ANGDLNVVVKTKQDGSKALLATYNAAGQNTYLTN 1990
Db 2424 IDGNTYTTATVTCNAWTFNIPVADIANFEATEEVVATVSD-----LAGNPATPATRNI 2477
Qy 1991 N---PAEAIDRINEQ---GIRFFHVND-----GHOE---PVVQGRNGID-----SSASG 2030
Db 2478 DTVAPAVTIDSISDDTGAQANDFITNDTDLVFNGETAEDSTVVVSLDGIETGVTANGAG 2537
Qy 2031 KHSVAIGFOAKAGAAVAIGRQTOAGN-----QSIATGDNNAQATGDSQIAI----- 2077
Db 2538 EWTLDYTGTLADGDYELSVTATNPNGNSATATQTIIVDVTAPTVTINAIAVDDIINAVE 2597
Qy 2078 -GTGNVVTGKHSGA-IGDPSTVKAD-NSYS---VGNNNQF-----IDATQ 2116
Db 2598 AGSPVAVSGTTGVEDGQVVTVIDGNTYATVGTNAWTFNIPVADIANFEATEEVVATV 2657
Qy 2117 TDVFG-----VGNNTVTES-----NSVALGSNSAISAGTHAGTQAKKSD 2156
Db 2658 SLAGNPATPATRNIITVDTTAPTVTINAIAVDDIINAVEAGSPVAVS-GTTTGV-----D 2712
Qy 2157 GTAGTTTGTAG--ATGTVKGFAGQTAAGVAVSGCAGCAERRIONVAAAGEVSATSTDAVNSQ 2214
Db 2713 GQVVTVIDGNTYATVGTNAWTFNIPVADIA-----NFEATEEVVATVSD-LAGNP 2763
Qy 2215 LYKATOGI-----ANATNELD 2230
Db 2764 ATPATRNIITVDTVAPAVNVELD 2784

RESULT 8

C48399
ABC-type transport protein ydbA.2 - Escherichia coli
C:Species: Escherichia coli
C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999
C:Accession: C48399; D64891; H64891
R:Moszer, I.; Glaser, P.; Danchin, A.
Biochimie 73, 1361-1374, 1991
A:Title: Multiple IS insertion sequences near the replication terminus in Escherichia
A:Reference number: A48399; MUID:92190338
A:Accession: C48399
A:Molecule type: DNA
A:Residues: 464-2020 <MOS>
A:Cross-references: GB:D85081; NID:g3041754
A:Experimental source: strain K-12
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:88089, NCBI:P:88090)
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-839, 'LDLPLYFOTSVIT' <BLA1>
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74483.1; PID:g17876
A:Experimental source: strain K-12, substrain M61655
A:Accession: H64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'M', 915-2020 <BLA2>
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74487.1; PID:g17876
A:Experimental source: strain K-12, substrain M61655
C:Genetics:
A:Gene: ydbA_2
A:Start codon: GTG

[illegible]

Db 494 -----SFHSSTTPT---TATGTGTATVSI 515
Qy 673 SN-----DIAKGVK-----GIDPINGLTTPKLTIVGSDKGTQLVIEQV-ASG----- 716
Db 516 SNITAPTEADGTIRTHGALDNGSGSIANGQDVSAQAQGLNAG--QIDIHOLNAKGSAPD 573
Qy 717 -----NDTKNIIRGLSFTPLPSITNAGGVRTTEBGNITISDE-----DKSAASIGDIL 764
Db 574 NHNGTIIIDAVHIIQAG-----SLNQNGNITTRQOLEIETDOLDNAHGKLSIAEADLA 627
Qy 765 NTGFNLKNSVSGVSTYNTVDFIDGNATTAKVTYDNTNOTSKVTYDVNVDEKTIELTG 824
Db 628 VSG--SLNQ--GEIATNQOIIHDGQOSTAVI--DWTNGTIGSGRVAQAQSLSNNG 681
Qy 825 DNGTKNIGVK-----TTTTTTNANGKATNFSTTDNDALVNAKDI---AENLNTIAK 874
Db 682 TLAADNKLIALQDDFYVERNIVAGNELSLSTRGSLKNSHTLOAGKRIRIKANNLDNAAQ 741
Qy 875 EIHTTKGTADTALQTFKVKGDATDEIITVGKGTQNGKVTNLTLLKGLNGLVATNKD 934
Db 742 GNIOGGTDTIGTQ--HNLNTRGLIDGQOTKI-QAGQMN--NIGTGRIYGDNIATAATRLD 797
Qy 935 GTVFTGINTQSLKAGDSTLNLKDGSLTKNPASNEIOVGADGVKFAKVDKNGSSTGIDG 994
Db 798 NQDENG--TGAIARAENLNL--GIGOLNRENLSLIYSGNDMAVGALDINGQATG--- 849
Qy 995 TSRTKDOIGFTGA-----NGSLDTRPHLTOKKLKVEVEITNTGI---NAGGK----- 1041
Db 850 ---KAQRHNAGATTIEAAGKMRGLVEKHLNTHNEHLKTOLVETGREHIVDYEAFCRHELL 905
Qy 1042 -----KITNIQSDITG-----NSND 1057
Db 906 REGTOHELGSVYNDESHLRTPPGAAHENHVKYDYEKVT--QKTQVQTAPAKIISND 963
Qy 1058 -AVTGRVYDLKTELESKINSAAKTAQNSLH-----EFSVADEQGN-HFTVSNPYSSYD 1109
Db 964 LTIDGKEVNTDSIIAGNLIIVQTEKDLGNEQTFGEKKFSENGKLSHWREKHKRD 1023
Qy 1110 TSKTSDVITFAGENGITTKVNGVVR-----VGIDQTKGUTP----- 1147
Db 1024 STGHSEQ--NYTLPEITRNIISIGSPAYESHKALSHHAPSOGTELPQSNGLISPLTSNSF 1082
Qy 1148 -----KLTGVNNGKGIVIDSK-----DGQNTI 1170
Db 1083 TPLESSSLIIPVNVKGYLVETDPRFANVROWLGSYMLDSLKDLPNNLHRLRGDGYEY 1142
Qy 1171 TGLSNTLANVTNDGAGHALSGLANDTKTRA-----ASIGDVLNA----- 1211
Db 1143 RLINQIAELT---GHRRLDGYQNDQEFKALMDNGATAARSMNLSVGIASAEQVAQL 1198
Qy 1212 -----GFILOQN----- 1218
Db 1199 TSDIWLVOKEVLPDGGTQTVLPQVYVRVKNKGIDGKALLSGSNPTQINVSGLKNSG 1258
Qy 1219 ---GEAVDFVSTYDTPVDIDGNATTAKVTYDDTSKTSKVYDVNVNDKTIETVSDKLGIV 1275
Db 1259 TIAGRNALIINT-DPLDNIGIRIHAQSAVATQDINNIGMISAE-QTLNLAGNINNS 1316
Qy 1276 KTTTLTKTSANGNATKFSAADQ-----DALVKASDIATHLNTLAGDI--QTAKGASO-- 1325
Db 1317 QSTTASSQNTQGSSTYLDRMAGIYITGKEGVLAQAQKADINIILAQISNQSEQOTRLQ 1376
Qy 1326 -----ASSASVYDADGNKVYIDSTDKKYYOVNDKQV---DKNKEVAKKLIV 1370
Db 1377 AGRDINLDTVQTSKHQATFPAD-NHVIRGSTNEVSGSIQTKGDVITLLSGNLLNAK---A 1432
Qy 1371 AOAOPTPDGTLA-----OMNVKSVINKEOVNDANKQGINEDNAFIKGLENAADTKTKNA 1425
Db 1433 AEVSSANGTLAVSAKNDINISAGINTHVDVDSKHTGRSGGNKLVITDKAQSHEHTAQS 1492
Qy 1426 AVTVG-----DLNVAQTPITTFAGDTGITAK-----KLGETLTIKGGQTDITNKILD 1471
Db 1493 STFEKQKVVLQAGNDANILGNSVIS---DNGTQIQAGNHVRIGTGT--QSQSEYTHQTQ 1547

Qy 1472 NNIGVAGTGDFTVKLAKDLNLNLSVNAAGGTRIDKGISFVDANGQAOKANTPVLASANGLD 1531
Db 1548 KS-GLMSAGIGFTIG-SKNTQENQ-----SOSNEHTGSTV---GSLKGDITIVA----- 1592
Qy 1532 LGGKRISNIGAAVDDNDAVN--FKQFNEVAKTVNNLNQNS-----NSCASLPP--VVTD 1582
Db 1593 --GKHYEIGTSTVSPEGNTIYAQSIDIAAHNKLNSNTTQTYEQKGLTVAFSPVDDL 1650
Qy 1583 NGKPLNGTGDGPRKAIKAGADGKYHANAN-GVPVDKDGKPTDADKLANLAHAKPLDAG 1641
Db 1651 AQAIAVAQSSQKVQOSKNDRVNAMAANAGWAQVOTGK---SAQNLANGTINAKOV--- 1704
Qy 1642 HOVVASLGGNSDAITLT-----NIKSTLPQIDTPMTGNANAGQAQSLPSLSAAQOSNAAS 1696
Db 1705 -----SISITYEQONRQTTQVQANQAQASQIQAGGKTTLLIATGAEEOSN--- 1749
Qy 1697 VKDVLNVCFLNQTNHQVDFVKAYDTVNFVNGTGADIT---SVRSADGTSMNIIVNTAL 1752
Db 1750 -----INAGSDVAGKAGTILIAD---NDITLQSAE 1777
Qy 1753 AATDDDGNVLIKAKDGKFKYKADDLMPNGSLKAGKSASDAKPTGLSL-VNPNAGRG-STG 1810
Db 1778 QSNTERGO-----NKSAGWNAGAASVFGQGSLSGVTAGGNVKGKNG 1821
Qy 1811 DAVALLNL-----SKAVFKSKDGTITTTVSSDGLSIOGKDNSSITLSKGLNVLGKVIS 1864
Db 1822 DSIITHRHSHIGDKGQSLTIQSGGDTTI-----KGAQVRGK-----GVQVNAKMLS 1866
Qy 1865 NVGKGTDKTDANVQOLNEVRNLLGLG-NAGND-----NADGNQ 1902
Db 1867 IQSVODRETYQSK--QONASAOQTVGVGFSAGGDYSQSKIRADHVSVTEQSGIYAGEDGYQ 1925
Qy 1903 VNIAIDIKDPNPGSSSNRTVIKAGTVLGGKGNNDPEKATGQVGVKDKGN-----A 1955
Db 1926 IKV-----GNHTDLKGIIIT-----STQSAEDKGNRFOTATIT 1959
Qy 1956 NGDLSNVWVKTKQDGSKKALLATYNAAGOT-----NYLTN----- 1990
Db 1960 HSDIKN---HSQYKESFGLGASASISGKTLQCGAQNKPONKHLTISVADKNSASSVGYG 2016
Qy 1991 NPAAEALDRINQGI--REFHVNDGNOEPVVOGRNGIDSSA-----SGKHSVAI-- 2036
Db 2017 SDSQSSSITKSGINTRNIQITDEAAQIRLTGKTAQKADIDTNTVTTDTERHSGSLKN 2076
Qy 2037 -----GFOAKADGEAAVA--ICRQTAQGNQSI-----AIGDNAQATGDQSIAGTGNV 2083
Db 2077 TFNKEAVQSELDLQRTVSQDFSKNVQQAQNTIENQHLDKADKEAAETAAAEALANGDWE 2136
Qy 2084 TGRHSGAIGDPSTVKADNSYSVGN-----NNQFIDATQTDVFGVG--NNITVTSNSVAL 2136
Db 2137 TAKRKAHEAQAADAAKADN-WOQGVILNMLASGLAAPTS---GAGIAAATASPAVSVAI 2192
Qy 2137 GSNSAISAGTHAGTQAKKSDGTAGTTTATAGTGTGKGFAGOTAV-GAVSVGAS----- 2188
Db 2193 GQHFKDLAQGNANKLTAQETAHVLAHVAAGVAAGVDNNALAGALSAGSSEAAAPYI 2252
Qy 2189 -----GAERRIQNVAAGEVSATSTDAVNGSOLYKATQCIANATNELDH 2231
Db 2253 SKWLYGKEKGSDLTAEEKETVTATINVLGTATGAAGVNSATDAAGSLNAQSAVEN 2308

RESULT 11

C83339
hypoetical protein PA2462 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83339
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A;Gene: PA2462

$\frac{1}{2}$
 $\frac{1}{3}$
 $\frac{1}{4}$
 $\frac{1}{5}$
 $\frac{1}{6}$
 $\frac{1}{7}$
 $\frac{1}{8}$
 $\frac{1}{9}$
 $\frac{1}{10}$
 $\frac{1}{11}$
 $\frac{1}{12}$
 $\frac{1}{13}$
 $\frac{1}{14}$
 $\frac{1}{15}$
 $\frac{1}{16}$
 $\frac{1}{17}$
 $\frac{1}{18}$
 $\frac{1}{19}$
 $\frac{1}{20}$
 $\frac{1}{21}$
 $\frac{1}{22}$
 $\frac{1}{23}$
 $\frac{1}{24}$
 $\frac{1}{25}$
 $\frac{1}{26}$
 $\frac{1}{27}$
 $\frac{1}{28}$
 $\frac{1}{29}$
 $\frac{1}{30}$
 $\frac{1}{31}$
 $\frac{1}{32}$
 $\frac{1}{33}$
 $\frac{1}{34}$
 $\frac{1}{35}$
 $\frac{1}{36}$
 $\frac{1}{37}$
 $\frac{1}{38}$
 $\frac{1}{39}$
 $\frac{1}{40}$
 $\frac{1}{41}$
 $\frac{1}{42}$
 $\frac{1}{43}$
 $\frac{1}{44}$
 $\frac{1}{45}$
 $\frac{1}{46}$
 $\frac{1}{47}$
 $\frac{1}{48}$
 $\frac{1}{49}$
 $\frac{1}{50}$
 $\frac{1}{51}$
 $\frac{1}{52}$
 $\frac{1}{53}$
 $\frac{1}{54}$
 $\frac{1}{55}$
 $\frac{1}{56}$
 $\frac{1}{57}$
 $\frac{1}{58}$
 $\frac{1}{59}$
 $\frac{1}{60}$
 $\frac{1}{61}$
 $\frac{1}{62}$
 $\frac{1}{63}$
 $\frac{1}{64}$
 $\frac{1}{65}$
 $\frac{1}{66}$
 $\frac{1}{67}$
 $\frac{1}{68}$
 $\frac{1}{69}$
 $\frac{1}{70}$
 $\frac{1}{71}$
 $\frac{1}{72}$
 $\frac{1}{73}$
 $\frac{1}{74}$
 $\frac{1}{75}$
 $\frac{1}{76}$
 $\frac{1}{77}$
 $\frac{1}{78}$
 $\frac{1}{79}$
 $\frac{1}{80}$
 $\frac{1}{81}$
 $\frac{1}{82}$
 $\frac{1}{83}$
 $\frac{1}{84}$
 $\frac{1}{85}$
 $\frac{1}{86}$
 $\frac{1}{87}$
 $\frac{1}{88}$
 $\frac{1}{89}$
 $\frac{1}{90}$
 $\frac{1}{91}$
 $\frac{1}{92}$
 $\frac{1}{93}$
 $\frac{1}{94}$
 $\frac{1}{95}$
 $\frac{1}{96}$
 $\frac{1}{97}$
 $\frac{1}{98}$
 $\frac{1}{99}$
 $\frac{1}{100}$

A;Gene: PA2462

1. $\frac{1}{2}$ 2. $\frac{1}{3}$ 3. $\frac{1}{4}$ 4. $\frac{1}{5}$ 5. $\frac{1}{6}$ 6. $\frac{1}{7}$ 7. $\frac{1}{8}$ 8. $\frac{1}{9}$ 9. $\frac{1}{10}$ 10. $\frac{1}{11}$ 11. $\frac{1}{12}$ 12. $\frac{1}{13}$ 13. $\frac{1}{14}$ 14. $\frac{1}{15}$ 15. $\frac{1}{16}$ 16. $\frac{1}{17}$ 17. $\frac{1}{18}$ 18. $\frac{1}{19}$ 19. $\frac{1}{20}$ 20. $\frac{1}{21}$ 21. $\frac{1}{22}$ 22. $\frac{1}{23}$ 23. $\frac{1}{24}$ 24. $\frac{1}{25}$ 25. $\frac{1}{26}$ 26. $\frac{1}{27}$ 27. $\frac{1}{28}$ 28. $\frac{1}{29}$ 29. $\frac{1}{30}$ 30. $\frac{1}{31}$ 31. $\frac{1}{32}$ 32. $\frac{1}{33}$ 33. $\frac{1}{34}$ 34. $\frac{1}{35}$ 35. $\frac{1}{36}$ 36. $\frac{1}{37}$ 37. $\frac{1}{38}$ 38. $\frac{1}{39}$ 39. $\frac{1}{40}$ 40. $\frac{1}{41}$ 41. $\frac{1}{42}$ 42. $\frac{1}{43}$ 43. $\frac{1}{44}$ 44. $\frac{1}{45}$ 45. $\frac{1}{46}$ 46. $\frac{1}{47}$ 47. $\frac{1}{48}$ 48. $\frac{1}{49}$ 49. $\frac{1}{50}$ 50. $\frac{1}{51}$ 51. $\frac{1}{52}$ 52. $\frac{1}{53}$ 53. $\frac{1}{54}$ 54. $\frac{1}{55}$ 55. $\frac{1}{56}$ 56. $\frac{1}{57}$ 57. $\frac{1}{58}$ 58. $\frac{1}{59}$ 59. $\frac{1}{60}$ 60. $\frac{1}{61}$ 61. $\frac{1}{62}$ 62. $\frac{1}{63}$ 63. $\frac{1}{64}$ 64. $\frac{1}{65}$ 65. $\frac{1}{66}$ 66. $\frac{1}{67}$ 67. $\frac{1}{68}$ 68. $\frac{1}{69}$ 69. $\frac{1}{70}$ 70. $\frac{1}{71}$ 71. $\frac{1}{72}$ 72. $\frac{1}{73}$ 73. $\frac{1}{74}$ 74. $\frac{1}{75}$ 75. $\frac{1}{76}$ 76. $\frac{1}{77}$ 77. $\frac{1}{78}$ 78. $\frac{1}{79}$ 79. $\frac{1}{80}$ 80. $\frac{1}{81}$ 81. $\frac{1}{82}$ 82. $\frac{1}{83}$ 83. $\frac{1}{84}$ 84. $\frac{1}{85}$ 85. $\frac{1}{86}$ 86. $\frac{1}{87}$ 87. $\frac{1}{88}$ 88. $\frac{1}{89}$ 89. $\frac{1}{90}$ 90. $\frac{1}{91}$ 91. $\frac{1}{92}$ 92. $\frac{1}{93}$ 93. $\frac{1}{94}$ 94. $\frac{1}{95}$ 95. $\frac{1}{96}$ 96. $\frac{1}{97}$ 97. $\frac{1}{98}$ 98. $\frac{1}{99}$ 99. $\frac{1}{100}$ 100. $\frac{1}{101}$ 101. $\frac{1}{102}$ 102. $\frac{1}{103}$ 103. $\frac{1}{104}$ 104. $\frac{1}{105}$ 105. $\frac{1}{106}$ 106. $\frac{1}{107}$ 107. $\frac{1}{108}$ 108. $\frac{1}{109}$ 109. $\frac{1}{110}$ 110. $\frac{1}{111}$ 111. $\frac{1}{112}$ 112. $\frac{1}{113}$ 113. $\frac{1}{114}$ 114. $\frac{1}{115}$ 115. $\frac{1}{116}$ 116. $\frac{1}{117}$ 117. $\frac{1}{118}$ 118. $\frac{1}{119}$ 119. $\frac{1}{120}$ 120. $\frac{1}{121}$ 121. $\frac{1}{122}$ 122. $\frac{1}{123}$ 123. $\frac{1}{124}$ 124. $\frac{1}{125}$ 125. $\frac{1}{126}$ 126. $\frac{1}{127}$ 127. $\frac{1}{128}$ 128. $\frac{1}{129}$ 129. $\frac{1}{130}$ 130. $\frac{1}{131}$ 131. $\frac{1}{132}$ 132. $\frac{1}{133}$ 133. $\frac{1}{134}$ 134. $\frac{1}{135}$ 135. $\frac{1}{136}$ 136. $\frac{1}{137}$ 137. $\frac{1}{138}$ 138. $\frac{1}{139}$ 139. $\frac{1}{140}$ 140. $\frac{1}{141}$ 141. $\frac{1}{142}$ 142. $\frac{1}{143}$ 143. $\frac{1}{144}$ 144. $\frac{1}{145}$ 145. $\frac{1}{146}$ 146. $\frac{1}{147}$ 147. $\frac{1}{148}$ 148. $\frac{1}{149}$ 149. $\frac{1}{150}$ 150. $\frac{1}{151}$ 151. $\frac{1}{152}$ 152. $\frac{1}{153}$ 153. $\frac{1}{154}$ 154. $\frac{1}{155}$ 155. $\frac{1}{156}$ 156. $\frac{1}{157}$ 157. $\frac{1}{158}$ 158. $\frac{1}{159}$ 159. $\frac{1}{160}$ 160. $\frac{1}{161}$ 161. $\frac{1}{162}$ 162. $\frac{1}{163}$ 163. $\frac{1}{164}$ 164. $\frac{1}{165}$ 165. $\frac{1}{166}$ 166. $\frac{1}{167}$ 167. $\frac{1}{168}$ 168. $\frac{1}{169}$ 169. $\frac{1}{170}$ 170. $\frac{1}{171}$ 171. $\frac{1}{172}$ 172. $\frac{1}{173}$ 173. $\frac{1}{174}$ 174. $\frac{1}{175}$ 175. $\frac{1}{176}$ 176. $\frac{1}{177}$ 177. $\frac{1}{178}$ 178. $\frac{1}{179}$ 179. $\frac{1}{180}$ 180. $\frac{1}{181}$ 181. $\frac{1}{182}$ 182. $\frac{1}{183}$ 183. $\frac{1}{184}$ 184. $\frac{1}{185}$ 185. $\frac{1}{186}$ 186. $\frac{1}{187}$ 187. $\frac{1}{188}$ 188. $\frac{1}{189}$ 189. $\frac{1}{190}$ 190. $\frac{1}{191}$ 191. $\frac{1}{192}$ 192. $\frac{1}{193}$ 193. $\frac{1}{194}$ 194. $\frac{1}{195}$ 195. $\frac{1}{196}$ 196. $\frac{1}{197}$ 197. $\frac{1}{198}$ 198. $\frac{1}{199}$ 199. $\frac{1}{200}$ 200. $\frac{1}{201}$ 201. $\frac{1}{202}$ 202. $\frac{1}{203}$ 203. $\frac{1}{204}$ 204. $\frac{1}{205}$ 205. $\frac{1}{206}$ 206. $\frac{1}{207}$ 207. $\frac{1}{208}$ 208. $\frac{1}{209}$ 209. $\frac{1}{210}$ 210. $\frac{1}{211}$ 211. $\frac{1}{212}$ 212. $\frac{1}{213}$ 213. $\frac{1}{214}$ 214. $\frac{1}{215}$ 215. $\frac{1}{216}$ 216. <

Db 1023 LSSQKIDVSLAGAFDNQ - - - AGSLDSRGFLTVKSAWLDNQGGTLSSAGALAVTSQG - - 1076

Db 1869 IG-----NOAGSISSOGRI.SVKSGRI.DNOGGA.VSSAGTIS.---ISSOGAI.NNOGCP 1916

1761 VLIRAKDGKFKYKADLL-MPNSGLKAGKSASDAKPTGLSLVNPAGKSGTGDVALNNLS 1819
 Db 1917 VVTDA--GAVLSASLNSOGGIVSAKGAEEIRGS---LN-NSQGGIGSGAGLALVA 1969
 Qy 1820 KAVPKSKDGTGTTTSSVSSDGIISIQKDNSS---TSLSKDGLNVGKVKLSNVGKGTDKTDA 1875
 Db 1970 DLVNSQNCRTAKGAIDA-NLKGIDQOGSLVSDTAIALDLRGGLVNVNRAOGLIATPG 2028
 Qy 1876 A-NVOOLNEVRNLLGLGNAGNDNADGNVNIAIDIKDPNSSSSNRVTIKAGTIVLGKG- 1933
 Db 2029 ALLUKQLGV-----DNSGGGEI-----SSDRSPTLAATLSNRGG 2064
 Qy 1934 ---NNDTEKL-----ATGGVQV-----GVDDKQNA 1955
 Db 2065 RVISGDSLTIRIAQALDNLQGLVLSAGSLDVAALVDFNHSIVASKGDTGTHGVNRLNE 2124
 Qy 1956 NG-----DLNVVWYTKQ-----DQSKKALATYNAAG-----QTNLYTNNAEALDR 1998
 Db 2125 AGRVVSEGAIDLTAQVSSAKGRIAAKGDQVTVGTTLQOGGELASQGTTLFD-ADSLDN 2183
 Qy 1999 INEQGIRFFHVDGN---OEPVVOGRNGIDSSASKHSVAI-----GFOAKADGEAAV 2048
 Db 2184 RGGGLVS---ADGGVTAEARQIDNRGG-EISSVAKVALAVREQLDNRGKGVIGDSELSL 2238
 Qy 2049 AIGRQTQAGNQSIAIGDNOAQTGDSIAIGTCNVVTKGH-----SGAIGDPS----- 2095
 Db 2239 TVQRLNQAQVLAASROGLHLDGAE-LLNGDGLLSQRLVDVTLSGALDNOGSGALVSE 2297
 Qy 2096 ---TVKADNSVSVGNQ---FIDATQ-----TDVPCVGNNTVTFESN 2132
 Db 2298 ESLTVKADQV-----NNOAGTFSSAGSLVLTSRGELNNOGRLVTDAGATLNSTGFDSNR 2352
 Qy 2133 SVALGSNSAISAGTHAGTQAKKSD--GTAGTITTAGAT----- 2168
 Db 2353 ACLYSAKCAVALRTGALNNSQKSGTGGTGTVLVAGLVNDNREGRISTRKGTLDANLKGLL 2412
 Qy 2169 ---GTVGFGAGQTA-----VQAVSVGASGAER---TONVAAGEVSA----- 2204
 Db 2413 QOGGSLVGERGVTLDLNGGTLDNHDLGLVSTPGALLRLQGLVNDNSVGEISSDRAFTL 2472
 Qy 2205 -----TSTDVANGSOLYKATOGIANATNLDHRHQNENKA 2240
 Db 2473 AANTLNNOGGRLLSEALTIRIAKTLNLSLKGQVL--ATYDGLATESQVLDNRAGTIGSKG 2530
 Qy 2241 NAGIS-SAMAMASMPQAYIPGRSMVTGGIATHNGQAV-----AYGLSKLSDNGQWY 2291
 Db 2531 DARISVTSILDNAEQSLVSEGRLELVADQVSNQGRIAARGVLEAAVG-TLLQOQGEIV 2589
 Qy 2292 FKINGSAD 2299
 Db 2590 SQ--GSLD 2595
 RESULT 12
 T03099
 N:Alternate names: pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000
 C:Accession: T03099; A40009; A28528; B29789
 R:Ekhardt, A.E.; Timpte, C.S.; Deluca, A.W.; Hill, R.L.
 J. Biol. Chem. 272, 33204-33210, 1997
 A:Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain
 A:Reference number: 214839; MUID:98070526
 A:Accession: T03099
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1328 <EC>
 A:Cross-references: EMBL:AF005273; NID:g2581863; PIDN:AAC62527.1; PID:g2581864
 R:Ekhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.
 J. Biol. Chem. 266, 9678-9686, 1991
 A:Title: Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in

A:Reference number: A40009; MUID:91236743
 A:Accession: A40009
 A:Molecule type: mRNA
 A:Residues: 12139-12167; 'T', 12169-13288 <EC>
 A:Cross-references: GB:M61883; NID:9454837; PIDN:AAA30998.1; PID:g164374
 R:Timpte, C.S.; Ekhardt, A.E.; Abernethy, J.L.; Hill, R.L.
 J. Biol. Chem. 263, 1081-1088, 1988
 A:Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical se
 A:Reference number: A28528; MUID:88087170
 A:Accession: A28528
 A:Molecule type: mRNA
 A:Residues: 12139-12167; 'T', 12169-12641 <TIM>
 A:Cross-references: GB:M21174; GB:J03512; NID:gl64321; PIDN:AAA30990.1; PID:g552360
 R:Ekhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Toumadje, A.; Johnson Jr., W.C.; Hil
 J. Biol. Chem. 262, 11339-11344, 1987
 A:Title: Structural properties of porcine submaxillary gland apomucin.
 A:Reference number: A92606; MUID:87280230
 A:Accession: B29789
 A:Molecule type: protein
 A:Residues: 1572-1607 <EC2>
 C:Superfamily: pig submaxillary mucin
 C:Keywords: tandem repeat

Query Match 4.5%; Score 520.5; DB 2; Length 13288;
 Best Local Similarity 20.9%; Pred. No. 2.8e-11;
 Matches 547; Conservative 316; Mismatches 1143; Indels 617; Gaps 114;

Qy 8 IFNKATCTFMVAACAKSHSGSSSTAGOVGSPVIRLTRVATLILVIGATLNGSAYA 67
 Db 1492 IYLOSTIGLSRSTEE---SGSTGISVIGSGKTSITLITVTSARS---SVAGSGTTGVTSV 1545
 Qy 68 QNNSKIAFGTTGNNDNASNEASIAIGSLAKAHANQAIIGGSKPDPR-NQAAQNKA 126
 Db 1546 -----ASGSGTSSGSGSPGATRASIG-----QKETRRISVAGSGGAPAVSSGAQA 1593
 Qy 127 HAKG-----KESIAIGGDVLAEGDASIAIGSDDLIDRSTNSKYPNGLLS 172
 Db 1594 SGAGPGTTASSVGVTTETARPVSAGSG---TTGTVSGASGS-----TGSSSGSPGATGA 1643
 Qy 173 TLIONHTVLRIDRSNOKYRRTAAEGHASTAVGAMAYAKGHFANAFGTSTABE- 227
 Db 1644 SIGQPETSRISVAGSSGA-----PAVSSGASQAAGTSCAGPGTTASSVGVTTETARPVS 1698
 Qy 228 ---GNVSLAVGLFAKAE-----KGYTTAIGSNQAQAINYGALALGADTRV 268
 Db 1699 SGTGTGVSGASGSGTSSGSGSPGATCASIGQPETSRISVAGSGGAPAVSSGA-SQAAGTS- 1756
 Qy 269 DLDYCIALGYCSQILNNNNNNNNKAYVPEGNCSNKKSKATGNGLFSIGSSTIKRKIINV 328
 Db 1757 -----GAGPGTTASSVGVTTETARPVS-AGSGTGTGTVSGASG---STGSSSGSPGATGA 1805
 Qy 329 GAGYEDTDAVNVA---QLKAVENLAKROIETKDDNGTCVKKKLGLETITIRKGETQADKL 385
 Db 1806 SIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGT-TASSVGVTT-----ETARPVS 1858
 Qy 386 TDNNNIGV-----TDNNTGLKVKLAKNLGLETIVSTKNTLTASEKVTVSGNNNTAEQLS 439
 Db 1859 AGSGTGTGVSGASGSGTSSGSGSPGATCASIGQPETSRISVAGSGGAPAVSSGAQAAGTS 1918
 Qy 440 G-----GLTFT---PTTNASTDKTVYTGDLKFTDN-----SNTALEDTTTRIT 479
 Db 1919 GAGPGTTASSVGVTTETARPVSAGSGTGTGTVSGASGSGTSSGSGSPGATCASIGQPETSRIS 1978
 Qy 480 -----KDKIGFSNKAQTVDENKPYLDKDKLVKCNSTLNNGL--TVNNTIGGSKNQI 529
 Db 1979 VAGSSGAPAVSSGASQAAGTSGAGP-----GTTASSVGVTTETARPVSAGSGTGTG 2027
 Qy 530 QV-GADGKIKFADVNVVNSNAKFGTTRITEBIGFADADGKVDKKSPLYLDKQLQVGGVK 588
 Db 2028 TVSGASG-----STGSSSGSPGATGASIG-----QPETSRSIS 2059

Qy	589	ITKDSG---INAGDQKISNVKDATDDDAVYKQLKQVQDADGALQSFSIREKQGEFT	645
Db	2060	VAGSGAPAVSSGASQAAGTSGAGPGCTTASSVGVTTETARPVSAGSGTGTVSGASGS--T	2117
Qy	646	ISNLYSNGNT-----PNTFETITFAGENG I--STSNIDIAKVKVKGIDPINGLJTLTPKLTV	698
Db	2118	GSSSGPGATGASIGOPET--SRISVAGSGAPAVSSGASQAAGTSGAGP--GTTASV--	2172
Qy	699	GSKDCKTQLVITBOVAGSNDTKNIIRGLSPTLPSITTNAGGVRITTEOQNTITSDDEKSKAA	758
Db	2173	----GVTEARPVA--GSGTTGTVSGASGSGTGSSGSPGATGASIGOPETS--RISVAG	2224
Qy	759	SIG--DILNTGCFNLKNSNSVGFVSNTVDFIDGNATTAKYVIDETNQTSKYVIDVNVDE	817
Db	2225	SSGAPAVSSGASQAAGTSGAGPGCTTASSV---GVTEARPVSAGSGTGTGTVS---GASG	2277
Qy	818	KTIELJTGDNKTN-KIG-VKTTTLLTTTANGKATNFSTTDNDALVNAKDAENLNTLAKE	875
Db	2278	STGSSGPGATGASIGOPETSRISSVAGSG-----APVSSGASQAAGT	2322
Qy	876	IHTTKGTADTALQTVKKDGDATDDDETIV---GKDGQTQNGKITVNTLKLXGENGLTVAT-	931
Db	2323	SGAGPGTTASSV-----GVTEARPVSAGSGTGTVSGASGSGTGSSGSPGATGASI	2374
Qy	932	NKDGTVTFGINTOSGLKAGDSTTLNKDGLSTKNPASNEQIOVGADGVKFAKVDKNSSTG	991
Db	2375	QOPETSRISSVAGSGAPAVSSGASQAAGTSGAGPGTTAS--SVGVTEARPVS-AGSGTGTG	2432
Qy	992	IDGTSRIKDOIQIFGANGSLDITTKPHLTKDKLKVGEVLTNTGINAGKKITNIQSG-	1049
Db	2433	TVSGASGSGTGSSGSPGATGA-----SIGOPETSRISSV-AGSGGAPAVSSGA	2478
Qy	1050	-----DITQNSNDAVTG---GRVYDLKTELESKINS-----AAKTAQ	1083
Db	2479	SQAAGTSGAGPGTTASSVGVTEARPVSAGSGTGTVSGASGSGTGSSGSPGATGASIGQ	2538
Qy	1084	NSLHEFSVADEQNHFTVSNPYSSYDTSKTSDVITFAGENGITTKVNKVVVRVIGIDTKG	1143
Db	2539	PETSRISSVAGSGAPAVSSGASQAAGTSGAGPGTT--ASSVGVTEARPVSAGSG-----	2591
Qy	1144	LTPPKLTVGNCKGIVIDSKQDNTITGLS-----NTLVNTDNGAGHALSGLANDTD	1198
Db	2592	-----TTGVSGASGSGTGSSGSPGATGASIGOPETSRISSVAGSGAPAVSSGASQAAG	2645
Qy	1199	K-----TRAASICDVLNAGFNLQNGEAVDPVSYDVFDFIDGN-----ATTAKV	1243
Db	2646	TSGAGPGTTASSVGVTEARPVSAGSG-----TTGTVSGASGSGTGSSGSPGATGASI	2698
Qy	1244	TYDDTSK-----TSKVYVDNVNDNKITIEVTSDKLGVKTTTLTK	1282
Db	2699	QOPETSRISSVAGSGAPAVSSGASQAAGTSGAGPGCTTASSVGVTEARPVSAGSGTGTGTV	2758
Qy	1283	TSANGNATKESAADGDALVKASDIATHLNTLAGDIQTAKGASQASSSASYVDADGNKVY	1342
Db	2759	SGASGSGTGSSGSPGATGASIGOPETSRISSVAG-----SSGAPAVSSGAS--QAAGTSGAG	2812
Qy	1343	DSIDKKYQYVNDKGQVDKNEVAKDLVAQAQDPDGTLAQMNKSVINKEOVNDANKQG	1402
Db	2813	PGTTASSVGVTE-----ETARPVSAGSGT--GT-----VSCASGSGT	2847
Qy	1403	INEDNAFIKLENAAKDITKNAAVTGD--LNAVAQTPITFAGDTGTTAKKLGETLTI	1459
Db	2848	SSSGSPGATGASIGOPETSRISSVAGSGAPAVSSGASQAAGTSGAGPGTTASSVGVTEA	2907
Qy	1460	KGGQDTNKLTDNNIGVWAGTDGFTVKLAKDLITNLNSVNWAGTTRIDEKGTSIFVDANCOAK	1519
Db	2908	RPSVAGS-----GTTGTVSGASGSGT-----SSSGSPGATGASIGOPETSRISSVAGSGG	2956
Qy	1520	ANTPVLISANGLDLIG-----GKRTSNICAAVDDNDVAFNEQFNEVAKTVNNLNQNSNG	1572
Db	2957	A--PAVSSGASQAAGTSGAGPGTTASSVGVTEETARPVS---AGSGTGTGTVSGASGSGTSS	3011
Qy	1573	ASLPFVVYTDANGRP-----INGTDKPKOAIKAGADGKYHYHANANGVYPVDKDGKPTDAD	1626

Db 3012 SGSPGATGASIGOPETSRISVAGSSGAP--AV--SSGASQAAGTSG-----ACPGGTAS 3061
 Qy 1627 KLANLAAGHGPLDAGHQVVASLGGNSDAITLTNLIKSTLP-----QIDPTNTGNANAGQAQ 1681
 Db 3062 SV-GVETARPVSAGSGTTGTTVSGASG---TCSSSGPGATGASIGOPETSRISVAGSS 3117
 Qy 1682 SLPSLS-----AAQQSNAASVKDVLNVGNPLQTNHNVDFVKAYDTVNVFNG----- 1728
 Db 3118 GAPAVSSGASQAAGTSGAGPGTTASSVGT-ETARPVSAGSGTTCTVSGASGSGTSSSGS 3176
 Qy 1729 ---TGADI-----TSVRSADG-----TMSNI-----IVNTALA 1753
 Db 3177 PGATGASIGOPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGTETARPSVA 3236
 Qy 1754 ATDDGNNVLKAKDGKFPYKADLLMPNGSLKACKSASDAKTPTGLSLVNPNAKG----GST 1809
 Db 3237 GSGTTGTV-----SGASGCTGSSSGPGATGASIGOPETSRISVAGSS 3279
 Qy 1810 GDAVALNLSKAVFKSKDGTFTTTVSSDGI-----STQK-----DNSSIT 1850
 Db 3280 GAPAVSSGASQAAGTSGAG-PGTTASSVGTETARPVSAGSGTTCTVSGASGSGTSSSGS 3338
 Qy 1951 LSKDGLNVGKVISNVG-KGTRDTDAAVQOQLNEVRNLLGLGNAGNDNADGNQVNIADIK 1909
 Db 3339 PGATGASIGOPETSRISVAGSSGAPAVS---SGASQAAGTSGAG-PGTTASSVGTETAR 3393
 Qy 1910 KDPNPGSSSNRTYKAGTVLGGKGNNDTEKLATG-GVQGVGDKDGNANGDLSNVWVKQ 1967
 Db 3394 RPSVAGSGT-----TGTVSGASGSGTSSSGPGATGASIG-----OPETSRISV-AG 3439
 Qy 1968 KDGSKKALLATYNAAGTNYLTNNPAEAIIDRINEOGRIFHFVNDGNQBPV--VOGRNGID 2025
 Db 3440 SSGAPAVSSGASQAAGTSGAGPGTTASSVGTETARPSVAGSGTTCTVSGASGSGTSS 3497
 Qy 2026 SSAGSKHISVAIG-----FQAKADGEAAVATGROTQAGNQSIATGDNAQTGDOSTAI 2078
 Db 3498 SGSPGATGASIGOPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGTETARP 3557
 Qy 2079 T--GNVYTGKHSAGIDPSPVTKADNSYVSGVNNNQFTDATQDFVCGVGNITVTESNSVAL 2136
 Db 3558 SVAGSGTTGTVSGASGSGTSS-----SSGSPGATGASIGOPETSRISVAGS-----SGAPAV 3608
 Qy 2137 GNSAISAGTHAGTQAKKSDGTAGTTTTA-----GATGTVKGFAGGTAVGAVSVGASG 2189
 Db 3609 SSGASQAAGT-SGAGPGTTASSVGTETARPVSAGSGTTGTTVSGASGSGTSSSGSPGATG 3667
 Qy 2190 AE-----RRIQ-----NVAAGE-----VSATSTDVANGSOLY 2216
 Db 3668 ASIGOPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGTETARPVSAGSGTT 3727
 Qy 2217 KATOGIANATNE-----LDHRIHONENK--ANAGISSAMAWAMPQAYIPGRSMVTGG 2267
 Db 3728 GTVSGASGSGTSSSGPGATGASIGOPETSRISVAGSSGAPAVSS-----GASQAAGT 3780
 Qy 2268 IATHNGOGAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGA 2310
 Db 3781 SGAGPGTTASSVGTETARP-----SVAGSG-TTGTIVSGASGS 3817
 RESULT 13
 A83412
 hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: A83412
 R:Stover, C.K.; Yuen, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lar-
 : Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opport-
 A:Reference number: A82950; MUID:20437337
 A:Accession: A83412

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091;
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874
NID:99947856; PIDN:AAG05263.1; GSPDB:GN001

Query Match 4.4%; Score 511; DB 2; Length 2468;
Best Local Similarity 20.9%; Pred. NO. 6.8e-12;
Matches 573; Conservative 292; Mismatches 1027; Indels 848; Gaps 125;

QY	12	ATGTFMAVACAKSHGGSSSSTAGQGVSSPVIRLVRVATLAIVLTGATLSNGSAYAQNNS	71
Dd	98	ADPLMAGYIQPESLAGFESLTGACVLGG-----MSAGTALLVCAAAIGAVSAISNS	149
QY	72	KIAFGTGGNN-----DNASASNE-----ASTAIGS-----LAK	99
Dd	150	SGGGGGSSVPDTPPPKAASGLKTAPDGSSISLGQAEGASVGIDTINGDGKPDLVTIAD	209
QY	100	AHAN-----QAIAI-----GGSKDPDRNOAAOKA-----GSHAK	129
Dd	210	ANGNFATPLNPPLTNGQTVVVTDPGNASPPAQVAPTDTTAPAPATDVQVAPDGSSVT	269
QY	130	GK-----ESTAIAGDVLAEGDASTAIGSDDLLDNRNSTNSKYPNGLLSTLIQHNVTLRQR	185
Dd	270	GKAEPSTGVGVDTGDCQPDTTVVVGPGGSF-----EVLNPPLTNGETVTWIV-----	318
QY	186	DSNGOKYRRTAAEHGHAHVAMAYAKAHFAFAFGTRSTAEGNTSLAVGLTAKAEKGYT	245
Dd	319	-----TDPAGNSTPVTVEAPDDTTAPATDVQVAPDGS-----SVTGAEPGAT	363
QY	246	IATGSNAQAINYCALGADPRVD---LDYGIALGYSQI---LNNNNNNNKAYV----P	296
Dd	364	V-----GVDTDGGOPTTVVVGPGGSFEVLPNPLTNGETVTWITDTP	407
QY	297	EGRSNIKSKATGNGLFSIGSSITKRKIINVAGVEDTDVAVNOLKAVENLAKROITF	356
Dd	408	AGNSTPVTAEA-----PDPPAPOVNASNGSVLSGTAAGVTII	446
QY	357	K-GDDNGTGVKKKLGETLIKGE---TQADKLTDNNNIGVV-----TDNN	398
Dd	447	VITDNG--NPIGQTSADANGNWSFTPEGSQLPDGTVVNVVARDAAGNSSPATSIITVDG	502
QY	399	TGLVKVLKANLSGLETVSTKNLTASEKVTVGSCN---NTAELOSGGLFTPTTNASTDKT	455
Dd	503	VAPNAPVPEPSNGSELSGTAEPGSSVLTLDGNGNPIGQTTADANGNWFFTPST-PLPDGT	561
QY	456	VYGTDLGKTTDNSNTALEDTTRI-----TKDKIGFSNKAGTVD--ENKPYLDDKDLXVG	507
Dd	562	VNVVVARDAAGNSPPASVTVDAVAPATPTVDPSNGCTTILSGTAEPGSSVLTLDGNGNPIG	621
QY	508	NSTLNNGG-----LTVNNTIGGSNKIOVGCADGIKADVNVVYSN----547	
Dd	622	QVTADSGSNWTTPSTPLPENGTVVWNATATDPGNSASSPASVTVDAVAPATPVVNPNGTT	681
QY	548	----AAKFCTRITE---BEIFADADGKVDKKSPLYLDKKQLQGVGGVKITKDSGINAGDO	600
Dd	682	LSGTABPGATVTLTDGNGNPIGOVTDAGS-----GNWSFTPTPLPNGVT	726
QY	601	KLSNVKDAYDSTDDA--VITYKOL---KVOQODADGALQSFRDEKGQEFTISNIYNSGN	654
Dd	727	VNATATDAGNSTSAGSSVTVDSVAPATVINPSNCT--TLSGTAEPGSSVLTLD--GNGN	782
QY	655	TPNTEFIIFAGENISISNDIAKGVKVGIDPINGLTTPKLTGVSDDKGKTLQVIEQVA	714
Dd	783	-PIGVOTADGSGCNWSFTPTPLADGTV-----VNATAT-----DPAGNTS-----G	822
QY	715	SGNDTKNIIRGLSPILPSITNAGGVRTTEQGNTWITSDEBKSKAASIGDILNTGFNLKNS	774
Dd	823	QGSTT---VDGVAPTTPTV-----NLSNGS	844

Qy	775	NSVGEVSTYNTVD	FDIGNATTAKVY	DETNOTSKVYDVNVDEKTI	ELTGDNGKTNKIGV	833																																															
Db	845	SLSGTAEPG	SVILTDGNG	-----NPIAEVADGS-	-----GN	876																																															
Qy	835	KTTVLTTTNANG	KATNFSTTD	--NDALVNAKDI	AELENLTLAKEIHTKCF--	ADTALQTF 890																																															
Db	877	WYTPSTPIANG	TVVNVVQA	DAGNSSPGASVTV	DSQAAPVVPNSGNTT	LSGTA----932																																															
Qy	891	KVKKGATDDE	ITVKGDTG	QNGKTVNTL	KLKE-----	NGLTV--ATNKDGTV 937																																															
Db	933	---EPGAT---	VTLTDG----	NGNPIQGV	TADSGNWSFTPG	TPLANGTVVNATASDPT- 981																																															
Qy	938	TFGINTOSGL	KAGDSTTLNK-	-DGLSTKNPASNEQI-	-----	QVG 974																																															
Db	982	-----GNTS	APASTTTVD	SVAPAAPVVP	NSGAIEISGTA	EPGATVTLTDGSGNPIGQVT 1033																																															
Qy	975	ADGVKFAK	VDKGN-----	SSTGIDGT--	SRITKDKQIGFT	GANGS---LDTTKP-----HLT 1022																																															
Db	1035	ADG-----	SGNWSFT	PPLADGT	VVNATADPAGNT	GGGSTVDAIAIATPTVNL 108																																															
Qy	1021	KDKLKVEI	ETNGINAGK	-----KITNIQ	SGDITQNSND	AVTGRVYDIKLTULESKI 1075																																															
Db	1088	NGSSLG	STAEPG	SVILTDG	NGNPIAEV	TADSGNNYTPSTPIANGTVVNVAQDAGN 114																																															
Qy	1076	NSAAKT	AQNSLHEFS	VADEOGN	HTVSNVSSY	DTSKTSD-----VITPAGENG- 112																																															
Db	1148	SSPPATV-----	IVDS	APAPV	NPNSGVV	ISGTAEGATVTLTDAGNPIQGVTA 119																																															
Qy	1126	-----	TTKVNKGV	VVRGIDQ	TGKLTTPK--	LVVGNNGKGI	VIDS	KDQNTITG 117																																													
Db	1200	DGSGNWS	FPTG	TPLANGTV	VATATDPT	GNTPQAA	TTTVD	VAVAPPV	IDPSNG-----125																																												
Qy	1173	LSN	TLANVTND	GAGHAL	SOGLAND	TOKTRAAS	IGDVLN	AGFNLOG	NEAVDPFVSTYD 123																																												
Db	1254	--TTIS	TACAGAKV	ILTDG	NGNPIG	ETTADSG	NWSFTPG	TPLANGTVVN	NAVO----130																																												
Qy	1233	FIDGN	ATTAKVY	DDTSKTSKVV-	-YDVNV	DNKTI	IEVTS	DKKLGV	KTTTLTKTSANGNAT 129																																												
Db	1308	PAGNTG	PGQSTTV	DVAVAPNT	PPVNP	NSGNLLN	CTAEPGS-	-----	TVTLTDGNGNPI 135																																												
Qy	1291	KFSAADG	D--ALV	KASDI--	ATHL	NLTAGD-----	IQAKAS	QAS	SSASVYDADGNK 133																																												
Db	1359	GQTTAD	SGNWSFT	PGSQPL	NGTVV	NVTASD	DAAGNTSL	PATTTVD	SSLPISQVDPDSNGS 141																																												
Qy	1340	VYD	STD	KKYQV	NDKQV	DNKEVAKD-----	KLVAQA	QTPDGT	LAQMNKVSINKIQV 139																																												
Db	1419	VIS	TADAGNT	IIITD	GNPN	IGQV	TADSGNWS	FTPG	IPLPDGTVVNVARS	PNVD 147																																											
Qy	1395	-----	NDANK	KQGIN	EDN	NAFI-----	KGLENA	AKD-----	TKTKN 142																																												
Db	1479	PAVIT	VDG	VAPAA	PV	IDPSNGTE	ISGTA	EAGATV	ILTDG	GGNPIQOATADSGNWTFTPS 153																																											
Qy	1425	AAVT	VG	D--L	NAVAQ	TPLTE	AGDTG	TTAKKL	GTLLTKG	QDNTKLTNNI	GVVAGT--D 148																																										
Db	1539	TP	LANGTV	IN	NAVAQD	--AGNT	SGPA-----	SVIT	DAI	APPAPV	NPNSGVV	ISGTAE 159																																									
Qy	1482	GFTV	KLAKOL	TLN	LSNV	NAGGTR	ID	EKGISF	VDANG	QAKANT	PTVLS	ANG	L	D	IGGKRIS	NIG 154																																					
Db	1591	GATV	----	ILTD	GNPN	IGQV	TADSGG--	NWSF	TPGT	PLANG	SVI	NAL	AQA	D	AAGN	SSPTS 164																																					
Qy	1542	AAV	D	ND	DAV	NFKQ	NEV	AKTV	VNN	LNQ	NSG	ASLP	VV	TV	D	ANG	R	ING	CTD	GK	P	KA	IKGA 160																														
Db	1646	ATVD-----	SL	AP	V	IDPS	NGSV	IAGT-----	166																																												
Qy	1602	DG	KY	Y	H	A	N	A	N	G	V	P	V	D	K	G	K	P	I--	T	D	A	D	K	L	I	A	N	I	A--	A	H	G	K	P	L	D	A	G--	H	O	V	A	S	L	G	N	S	D	A	I 165		
Db	1670	----	AE	A	G	A	T	V	I	L	T	D	G	N	P	I	Q	O	V	T	A	D	S	G	N	W	S	F	T	P	G	T	P	L	S	N	G	T	V	V	N	A	V	A	D	A	A	G	N	T	S	G	P 172
Qy	1656	TL	T	N	I	K	S	T	L	P--	Q	I	D	P	T	P	T	G	N	A	G	O	A	Q	S	L	P	S	I	A	A	O	Q	S	N	A	S	K	V	D	L	N	V	G	F	N	L	O	T	H	N	Q 171	
Db	1726	V	S	T	T	D	A	V	A	P	A	P	V	I	D	P	S	N	G	V	E	L	S	G	T	A	E-----	P	G	R	V	I	L	T	O	G-----	NG	N	P 176														
Qy	1714	V	D	F	V	K	A	I	D	T	V	NF-----	V	N	G	T	A																																				

Db 1768 IGTGLADGSGNWSFTCTPCTPLANGTVVNAVAQDPAGNMSGPASTVDTVPATVPINPSNG 1827
QY 1760 NVL-----IKAK-----DGFKYKADLMPNGSLKAGKSASDAKPTGLSLVNPNGAGKSGTG 1810
Db 1828 SVITGTAEGVAKVILTDG-----NGNIGETTAGDSGNWTFP-GTPLAN-----GTVI 1875
QY 1811 DAVALNNLSKAVPKSGDGTITTV-----SSDG-----ISIQCK 1844
Db 1876 NAVAEDAAGNA-----SGPASTTVDSVAPSAPLLSISADGALLTGAEPNSQVRIVVNGD 1930
QY 1845 DNSSITLSKDL-----NVGKVISN----- 1865
Db 1931 TANPIITVVDGAGNLSFPAPPLITGELIAGVADAGNVSGPATINADPLAPPTISVPE 1990
QY 1866 -----VGRGKGT-DTDAANVQQLNEVRNLLGLGNAG-----NDNADGNQV 1903
Db 1991 AADTWINAAEIGDGIQDVTVRTMQGVQVTVKFAQONGYAEVSHLTAGDIAAGNLT 2050
QY 1904 -----NTADIKKDNSSGSSN-----RTVIKAGTVLGGKN-----ND 1936
Db 2051 LTLTPPGMGFPFEGASTVADINGGTASTPVPFTIDTTPATPVLVLGNILTISAEPG 2110
QY 1937 TEKLAT-----GGV-----QVGVDKDNANGDLSNVVTKQDGSKKALLATYNAAGQNYLT 1989
Db 2111 TELTVTVDVGGVATATVTTADNSGLAS-----LNLTLTDLDIDFSWDQLL-----NA--QVSVVG 2162
QY 1990 NPAAEADIRNEGIGIRFFHVNDGNQPVVQGRNGIDSSAGSKHSVAIGFOAKADGAAVA 2049
Db 2163 RDPAG-----NPSNTASIGVGTSEIQPVITIGNEGLDVLNPLNP-RFGSGTTEPSSVV 2216
QY 2050 IGRQTAGNOSTAIGNOAQATQOSTAIGTVGNVTVKHSIGAIGDPSTVRADNSYSVGNNN 2109
Db 2217 IRVITPALNVEL-LPIQADSSGNSFNLNLSPTILT-----QLGLN- 2255
QY 2110 QFIDATQOTDVGVCN-----NITVTESNVALGNSAISA-----GTHAGTQAKKSDGTA 2159
Db 2256 -----ITDLNLGSQLSEFLVSTDSN-----GNDSAAYGTLTPNGLSLMIGQIDVNGTS 2305
QY 2160 GTTTTAGATG-----VRGFAQATGAVGVASGASGAERRIQNVAAE-----VSATSTDAV-- 2210
Db 2306 GDDVLSGANGSSHERINGGDSLDLIFNVGTG-----DHSVAGNGNDTIQITATDFVSI 2357
QY 2211 -----NGSOLYKATQGIATNATNLDHRIHONENKANAGISSAMAMASMPQAYI 2258
Db 2358 DGGAGDFTLVLANGIDLDYNAVGVGTLSNL--ERI--DLGKGDGSGVLTLTAAEVDAITD 2413
QY 2259 PGRSM-VTG-GIATHNGOGAVAVCLSKLSLSDNGQWVFKING 2296
Db 2414 ANNTLQITGENNDTLNVVGAVENTGTQ-----ING 2444

RESULT 14
E85822
probable invasin 23135 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <STO>
A:Cross-references: GB:AE005174; NID:gl2516151; PIDN:AA057041.1; GSPDB:GN00145; UWGP:231
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 23135

Query Match 4.3%; Score 504.5; DB 2; Length 2660;
Best Local Similarity 20.2%; Pred. No. 1.3e-11;
Matches 550; Conservative 329; Mismatches 979; Indels 869; Gaps 138;

QY 83 NASASNEASTA-----IGSLAKAHANQAIAGGSKPPRNOAANOKAG---SHAKGESI 134
Db 132 NSSGNLEQQIASTSQIGSLAEDMS-----EQANMARGWASSOASGAMT- 178
QY 135 AIGGDVLAE-GDASIAIGSDDLYLDRNSTNSKY-----PGLLSTLIQNHITVLR-- 182
Db 179 ----DMLSRGRTARITLGVDE--DFSLKNSQDFLHPWYETDNL--FSSQHTLHRD 228
QY 183 ---QIRDSNGSQRYRTAABG-----HASTAVGA-----MAYAKGHFANAFG 221
Db 229 ERTQINNGLRHFTPTWMSGINFFFDHLSRHSRAGIAGYWRDYLKLSNGNRLTN 288
QY 222 TRSTAE--GNYSL--AVGLTAKAE-----KGYTIAI-----GSSNAQ 253
Db 289 WRSAPELDNDYEAPANGWDVRAEGLPAPHLGGKLVLYEQYCYDEVALFDKDDROSNPH 348
QY 254 AI-----NYGALALGA-----DTRVOLDY-----GIAL----- 276
Db 349 AITAGLNYTPFLMTFSAEQRCQKQGENDTFRAVDFTWQPSAMQKOLDPNEVDARSLA 408
QY 277 GYGSQILNNNN-----NNNKAYV-----PEGNSNIKSKKATGNGL--FSGSST 320
Db 409 GSRFDLVDRNNNIVLEVRKKELVRLTLDPTVGKSGEVKSLVSSLOTKYALKGYNVEATA 468
QY 321 IKR--KIINVAGY-----EDTDVNVVAQLKAVENLAKRQ-----I 354
Db 469 LEAAGKVVVTGDKILVTLPAYRFTSTPETDNTWPIEVTAEYKGFNSREQSMVVVOAP 528
QY 355 TFRGDDONGTGVKKLGETLTIKGETQ-----ADKLTDDNNIGVVTDDNTGLKVKLAKNL 409
Db 529 TLSQKDSVSLS---SOTLSADSHSTATLTFIAHDAAGNPVIGLV-----LSTRH 575
QY 410 SGLTSTVKMLTASEKVVTVGSGNNNAEL-----OSGGLTFTPTTNASPTKVTVYGDGLK-- 463
Db 576 EGVO-----DITLSDMKNDGDSVYQIILTTGAMSGTLTLPQLN-----GYDAAKAP 622
QY 464 ----FTDNSNTALEDITRITKDKIGFSNKAGTV--DEN-KPYLDK-----DKL 504
Db 623 AVVNIILSVSSRTHSSITKIDKRYLSGNPIEVTVELRDENDKPKVEKQKQOLNTAVSIDNV 682
QY 505 KVG-----NSTLNNNGGLTVN-----NT 521
Db 683 KPGVTTDKETADGVYKATYATYKSGSLTAKLLMQWNEDLHTAGFIIDANPQSAKIAT 742
QY 522 IGSNKQIOVQADGIKPADVNVNVNNAKFGTTRITEEIEGFADADGV----- 570
Db 743 LSASNNGLVLANENAANTVSVNV-----ADEGSNPINDHTVTFVAVLSGSATSFNNQNTAKT 797
QY 571 -----DKSPYLDKKQLQGVGVKITKDSGIN-----AGDKISNVKDATDDTDAV 615
Db 798 DVNGLATFDUKS-----SKQ-EDNTVEVTLENGKQTLIVSVGDSSTAQVD----- 843
QY 616 TYKQLKQVQODA-DGALQSFISIRDEKGOEFTISNLYSNGNTPNTFTITFAGENGISIN 674
Db 844 LOKSKEVNVADGDSATMTATVRDAG-----NLL-----NDVKVTF 880
QY 675 DIAKGVKVGIDPIN---GLTTPKLTVGSQDKGTQLVIEQVAGSNDTKNII----- 723
Db 881 NVNSAAAKLSQTEVNSHDGIATATLT--SLKNG-DYTVTASVSSGSOANQOVFIGDQST 937
QY 724 RGLSPTLPS-----ITNAGGVRTT-----EOGNTITSDEKSKAASIGDILNTGFNLKN-- 772
Db 938 AALTLSVPFGDITVTNTAPLHMTATLQDKNGNPL---KDKIEITFSPNDVASRFSINSNG 994
QY 773 ----NNSNVGFVSTYNTVDFIDGNATTAKVYDETNQTSKVTVYDVNVVDYIETLGTGNKG 828
Db 995 KGMTDSNGTAISLTGTL--AGTHMITARLANSNVSDTPQTFVADKDRVVVLOTSKAE 1052
QY 829 TNKIGVKTTLTTTNNANGKATNFSTTDNDALVNAKO-----IAENLNTLAKEIHTTKGTAD 884

Db 1053 IIGNGVDETLTAT-----VKDPEDNVVKNLSVYFR-----TSPAD 1088
Qy 885 TALQTFKVKKGADGADDEITVKGDTQNGKTVNTLKLKGENGLVATNKDGTVFINGIQ 944
Db 1089 TQL-----SLNARTN-----ENGIAEVLTK-GTV-LGVHTA 1118
Qy 945 SGLKAGDSTTLNKDG--LSIKNPASNEQIQV-----GADGVKFAKVDKGNSTGTI 992
Db 1119 EAILNG-----NRDTKIVNIAPDASNAQVTLNIPAAQVVTNNSDVQLTATVDPSPHPV 1174
Qy 993 DG-TSRIKDKQ---IGFTGANGSLDTTKPHLTOKLKYGEVEITNTGINAGKKITNIQS 1048
Db 1175 AGITVNFMPQDVAANFTLENNGIAITQAN-----GEAHVTLKGGKAGCTHTVATL 1226
Qy 1049 GDITONSNDATGGRVYDLKTELESKINSAKTAONSLEEFSEVADQGNHFTVSNPYSSY 1108
Db 1227 NN---NASDAQPVTEVADKDSAVVVLQTSKAEIINGVDETTLT-----ATVKDP---F 1274
Qy 1109 DTSKTSDVITFA-----GENGITTKVNGKVVVRVGDIDQTKGLTTPKLTVGN 1153
Db 1275 DNAVKDLQVTFSTNPADPQLSQSKSNTNSGVAEVTFKGTV-LGV-HTAEATLPN---GN 1329
Qy 1154 NNGK--GVIDSKQDNT-----ITGLSNT--LANVTDNGAGHALSQGLANDTDKTR 1201
Db 1330 NDTKIVNIAPDASNAQVTLNIPAAQVVTNNSDVQLTATVDPSPHPVAGITVNFMPQD 1389
Qy 1202 AASIGDVLNAGFNLOGNGEAVDFVSTYDVFIDGNA---TTAKVTYDDTFSKTSKVYVD 1257
Db 1390 VA-----ANFTLENNGIAITQANGEAHVTLKGGKAGCTHTVATLSNNNTSDSQPVTF- 1441
Qy 1258 VNVDNKTI-----EVTSDKGLGVKTTTLTKT-----SANGNATKFSAD----- 1296
Db 1442 --VADTKSALVQLQTSKNEITGN--GVDSATLTLTATVKQDFQNEVNLPVTFSTASSGLT 1496
Qy 1297 ---GDALVKASDIATHLNTLAGDIOFTAKGASQASSASYVDADGNKVIYDSTDKKYQVN 1353
Db 1497 LTPGESNTNESGIAQ--ATLAG---VAFGEQTVTASLANNGASDNKTVHFTGD----- 1544
Qy 1354 DRGOVDKNKEVAKDLVAAQAPDPDGTLA---QMNVKYSINKEVQVNDANKKOGINEDNAF- 1409
Db 1545 -----TAAAKIIEITPVPSDSIIAGTPQNSSGVSITATVV-----DNGGFP 1584
Qy 1410 IKGL-----ENAKDTKYKNAATVGDLDNNAVAQTPLTAGDTGTTAKKLGETLTIKGQT 1464
Db 1585 VRGVTVNFETSAATAEWTNGCOAVTNEOGXATVYVTNTRSSIESGARDPTVEASLENGSS 1644
Qy 1465 DTNKLTDNNIGVVAGTDGFTV-----KLAKDLTNL-----NSVNAGGTRIDEXGISFVD 1513
Db 1645 TLS--TSINVNADASTAHLTLQALFDPVSAGDTTNLIEVKDNGYNGVPOQEVTLVSP 1702
Qy 1514 ANGOAKANTPVLSAN-----GLDLGOKRISNIGAAVDDDDAVNFQKOFNEVAKTVNNLN 1566
Db 1703 SEGVTPSNNAITVTHDGNFYASFETAXRAGVYQVVTATLENXDSMQ-----QTVTVVP 1754
Qy 1567 NGSNGASLFPVFTDANGKPI--NGTDGKPKQA--IKGADGKYYHANANG-----VPVD-- 1616
Db 1755 NVANAPISLA-----ASKDPVTANNNDLTLTATVADTEG-----NAIANSEVFTTLPDVR 1806
Qy 1617 -----KDGKPIITDADKLANLAHAKPLDAGHQVVASL--GGNSDAITLTNIKSTLPQDIT 1669
Db 1807 ANFTLGDGKVVYDTDEGKAKVTLKGTKAGA-HTVVTASMAGGKSEQLVVNFIAIDL----- 1860
Qy 1670 PNTGNANAOASLPSLSAAQOSNASVSKDVLNVFNLTQTNHNVQDFVKAYDTVNFVNG- 1728
Db 1861 --TAQVN-----LNVTEDNFIA-----NNVGMTRLQATVTDGNGN 1893
Qy 1729 --TGADITSVRSADGTMSTNITVNTALATDDGN--VLIKAKDGKFKYKADDLMPNGSLK 1783
Db 1894 PLANEAVTTLTPADVASSTFLQGGGSAITDINGKAEVTLGSKGTPVTVSVNNYV--- 1950
Qy 1784 AGKSADAKTPTGLSLNPNPNAKGSTGDAVALNLSKAVFKSKDGTITTTTSSD----- 1837

Db 1951 ----VSDTKQVTLIA----DAG----TAKLASLTSVYSFVSVTTEGATMTASVTDANGNPV 1999
Qy 1838 -GISIOCKDNSSITLSKGLNVGK-----VISNVGKTGTDQDAANVQQLNEVRNLLCLG 1891
Db 2000 EGIAKNFR-GTSVTLSTSVETDDRGFAEILVTSFVGLKTVSASLADKPTFV-----IS 2053
Qy 1892 NAGNDNADQGVNIADIKDPNSGSSSNRTVIKAG-----T 1927
Db 2054 RLLNAKADINSATITSLLEIPGOVVMVAQDVAVKARVNDQFQGNPILNESVTFSAEPPEHMT 2113
Qy 1928 VLGGKGNNDTEKLATGGVQGVQVDKDN-----ANGDLSNVVVKTKD-----G 1970
Db 2114 ISQNVSTDTHGIAE--VTMTPERNSGYMKVKSALANGS-----SYEKDLVVIDQKLTLS 2165
Qy 1971 SKKALLATYNAAGOTNVLTNPAEADIRNEQGIKF-----FHVNDGNOEPV 2017
Db 2166 ASSPLIGVNSPTGATLTATLTSACTP-VEGOVINESVTPPEGATLSGGKVRTNSSGOAPV 2224
Qy 2018 VQGRNGIDSSASGKHISVAIGFOAKADGEAAVAIGRQTOAGNQSTAIGDQAQNTGDSIAI 2077
Db 2225 V-----LTSNKGVTYTVTASFH-----NGVTIQOTIVK 2253
Qy 2078 GTGNVVTGKHSAGIDPSTYKADNSYSVGNNNQFIDATOTDVEGVGN---NITV---TES 2131
Db 2254 VTGNSSTAHVASFADPSTIAATNS-----DLSTLKATFED--GSGNLEGLTVYFALKS 2306
Qy 2132 NSVALGNSAISAGTHAGTQAKKSDGTAGTTTATGATCTVKGFAGQTAVGAVSVG---AS 2188
Db 2307 GSATLTSLTAVT-----DQNGIATTSVRCALTGSVTVSAVTTAGGMQTVDTILVA 2356
Qy 2189 GAERRIQNVAAGEVSATSTDAVNGSOLYKATQGTANATNELDRIHONENKANAGISSAM 2248
Db 2357 GPADASOSVLKANRRSLKGLDFTSAELHLVL-----HDSGNPIKVSLEPEVQ 2405
Qy 2249 AVASMPQAYIPG-----RSWVTG---GIAT-----HNGQCAVAVGLSKLD 2286
Db 2406 SCTNAPYQVVSADIDSKNFSGEYKATYTGGEGLATLPLVNGVHQAGLSTTIQFTRAED 2465
Qy 2287 N-GOMVFKING-----SADTQGHVGA 2306
Db 2466 KIMSGTVLVNGANLPTTTPPSQGTGA 2492

RESULT 15

H81193
hemagglutinin/hemolysin-related protein NMB0493 [Imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81193
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755
A:Accession: H81193
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2703 <FET>
A:Cross-references: GB:AB002405; GB:AB002098; NID:g7225708; PIDN:AAF40927.1; PID:g722 A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0493

Query Match 4.3%; Score 504.5; DB 2; Length 2703;
Best Local Similarity 20.4%; Pred. No. 1.4e-11;
Matches 557; Conservative 337; Mismatches 1044; Indels 791; Gaps 136;
Qy 1 MNH-IYKVIKFKATGTFMAVAPCAKSHSGSSST-----AGQVGSFVTRL 46
Db 1 MNRTLYKVFENKHNRCMIATAENAK-REGKNTADTQAVGILPNDIAGPAGFIHSISVSF 59


```

Db 1966 GSTTHRHVGS-----TGKTTIRSGDITLKGVLIGKGIQADTRNLHIE 2012
QY 1870 -TKDDAANVOQLN-EVRNLLGLGNAGN-----DNA-----DGNQVNI 1905
Db 2013 SVQDTETYQSKQNGNVQTVGCGFSASGSYRQSKVKADHASVTCOSGIYAGEDQYIKV 2072
QY 1906 ADIKDPNSGSSNRVTIKAGTVLGG-----KGNN--DTEKLATGGVQVGVKDGKNANG- 1957
Db 2073 RD-----NTDLKGGIITSSQSAEDKGNLFOTATLTASDIQNHRSRYEGRSFGI 2120
QY 1958 ----DLSNVW--VKYQKDG---SKKALLATYNAAGOTNYLTNNPAFAIDRINEQGIREFH 2008
Db 2121 GGSFDLNGGWDGTVTDKQGRPTDRISPAAGYSGDGSKNSTTRSG-----VNTHNI---H 2172
QY 2009 VNDGNQEPVVOGR-----NGIDSSASGKHSVAIGFOAKADGEAAVA--IGRQTQA 2056
Db 2173 LTDEAQLARTGRTAKETEAREIYTCIDTETADQHS---GHLKNSFDKDAVAKEINLQREV 2229
QY 2057 GNQSTAIQDQAQATGQSTAICTGNVVTCKHSGAI-----GDPSTVKADNSYSVG 2106
Db 2230 TRE---FGRNA-AQAAVAADVADKLGNTOQSYERYQEARTLLEAELOQNTDSEAEKAAPRASLG 2285
QY 2107 NNNQFIDATQT--DVF---GVCNNITVTESNVAIGSNATSAGTHAGTQAKKSDGTAGT 2161
Db 2286 QVNAYLAENQSRDYTWKEGGIGRSILHGAAGCLTTGSLGGILAGGTSIAAPYLDKAAEN 2345
QY 2162 TTTAGATGTVKGFAGOTAVGAVSVG-ASGAERRIONVAAGEVSATSTDAVNGSOLYKATQ 2220
Db 2346 LGPACK-----AARNALGGAAGIYATCG-----SGGAVVGANVD-WNNRQLHPKEM 2390
QY 2221 GIANATNE-LDHRHONENKANAGISSAM 2248
Db 2391 ALADKYAEALKREVEKREGKTISSQEAAM 2419

```

Search completed: September 13, 2001, 12:42:28
Job time: 32814 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 12:53:37 ; Search time 47.77 Seconds
(without alignments)
1659.351 Million cell updates/sec

Title: US-09-361-619-11

Perfect score: 11694

Sequence: 1 MNHLYKVFKNKATGTFMAVA.....NGSADTQGHVGAAGAGFHF 2314

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 'SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	642.5	5.5	2249	1 190K_RICRI	P15921 rickettsia
2	556.5	4.8	2003	1 YDBA_ECOLI	P33666 escherichia
3	492.5	4.2	1656	1 OMPB_RICJA	O08653 r outer mem
4	450.5	3.9	1577	1 HLYA_PROMI	P16466 proteus mir
5	448	3.8	3591	1 FIAB_BORPE	P12255 bordetella
6	445.5	3.8	1643	1 OMPB_RICPR	Q53020 r outer mem
7	444.5	3.8	1654	1 OMPB_RICRI	Q53047 r outer mem
8	438.5	3.7	1645	1 OMPB_RICTY	P65989 r outer mem
9	407.5	3.5	1902	1 P2P_LACPA	Q02470 lactobacilli
10	406	3.5	1848	1 CBPA_CLOCL	P38058 clostridium
11	404	3.5	1608	1 HLYA_SERMA	P15320 serratia ma
12	403	3.4	2334	1 WAPA_BACSU	Q07833 bacillus su
13	392.5	3.4	1902	1 P1P_LACLC	P16271 lactococcus
14	392.5	3.4	1902	1 P2P_LACLA	P15293 lactococcus
15	379	3.2	1183	1 CNA_STAAU	Q53654 staphylococ
16	379	3.2	1569	1 YFJA_ECOLI	P52143 escherichia
17	379	3.2	1902	1 P3P_LACLC	P15292 lactococcus
18	376	3.2	1325	1 YDEK_ECOLI	P32051 escherichia
19	362.5	3.1	1300	1 120K_RICRI	P14914 rickettsia
20	356	3.0	1829	1 FRPC_NEIMC	P55127 nelisseria m
21	349	3.0	1567	1 ICEN_XANTC	P18127 xanthomonas
22	336.5	2.9	1694	1 IGA0_HAEIN	P44969 haemophilus
23	336	2.9	2628	1 HAGA_FORGI	Q51845 porphyromon
24	335	2.9	1286	1 AIDA_ECOLI	Q03155 escherichia
25	333	2.8	1829	1 FRPC_NEIMB	Q91yv5 nelisseria m
26	331.5	2.8	1702	1 IGA2_HAEIN	P45384 haemophilus
27	320.5	2.7	1025	1 SLAP_CAUCR	P35828 caulobacter
28	311.5	2.7	1176	1 SLAP_BACSH	P38537 bacillus sp
29	311	2.7	1228	1 SLAP_BACST	P35825 bacillus st
30	310.5	2.7	1597	1 GTF1_STRDO	P11001 streptococc
31	309.5	2.6	1039	1 AG43_ECOLI	P39180 escherichia
32	309	2.6	2491	1 TALA_DICDI	P54633 dictyosteli
33	308.5	2.6	1849	1 IGA4_HAEIN	P45386 haemophilus

RESULT 1

190K_RICRI	1	TOXA_CLODI	2710	1	P16154 clostridium
34	307.5	2.6	918	1	P34487 caenorhabdi
35	302.5	2.6	918	1	P20469 pantoea ana
36	302.5	2.6	1322	1	P27470 streptococ
37	296.5	2.5	1592	1	P38536 t amylpull
38	287.5	2.5	1861	1	Q90953 gallus gall
39	286	2.4	3562	1	P16239 erwinia her
40	284	2.4	1258	1	Q10411 schizosacch
41	283	2.4	1957	1	P35827 campylobact
42	281	2.4	933	1	P32334 saccharomyc
43	275.5	2.4	1306	1	P49331 streptococ
44	275	2.4	1462	1	O33479 pseudomonas
45	272.5	2.3	1196	1	

ALIGNMENTS

190K_RICRI STANDARD; PRT; 2249 AA.

AC P15921;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-OCT-1994 (Rel. 30, Last annotation update)

DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).

OS Rickettsia rickettsii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=783;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R;

RX MEDLINE=90354033; PubMed=2117568;

RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;

RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";

RL Infect. Immun. 58:2760-2769(1990).

CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.

CC -!- PTM: GLYCOSYLATED (POSSIBLE).

CC -!- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M31227; AAA26380.1; .

PIR; A41477; A41477.

Antigen; Repeat; Signal; S-layer; Glycoprotein.

FT SIGNAL 1 28 POTENTIAL.

FT CHAIN 29 2249 130 KDA ANTIGEN.

FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.

FT REPEAT 212 286 A (TYPE I).

FT REPEAT 287 358 B (TYPE II).

FT REPEAT 359 430 C (TYPE II).

FT REPEAT 431 505 D (TYPE I).

FT REPEAT 506 577 E (TYPE II).

FT REPEAT 578 652 F (TYPE I).

FT REPEAT 653 724 G (TYPE II).

FT REPEAT 725 799 H (TYPE I).

FT REPEAT 800 874 I (TYPE I).

FT REPEAT 875 949 J (TYPE I).

FT REPEAT 950 1021 K (TYPE II).

FT REPEAT 1022 1093 L (TYPE II).

FT REPEAT 1094 1165 M (TYPE II).

FT REPEAT 1166 1180 TYPE I (INCOMPLETE).

SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match		5.5%	Score 642.5;	DB 1;	Length 2249;
Best Local Similarity		23.6%	Pred. No. 7.0e-16;		
Matches		522;	Conservative	210;	Mismatches 776; Indels 707; Gaps 114;
Qy	261	ALGADTRVDDYGTALGYSOILANNNNNNKAVVPENGSGNIKSKATGNGLESISST	320		
Db	36	ALGVATG-----IATNNAAFSNVGNMNEITAAAGVANGTPAGGPQNNAFTYGGD-	89		
Qy	321	IKRKIIINVAGYEDTDVAVNA-----OLKAIVENLAKRQITFKGDDNGTVGKKLGEFTLI	375		
Db	90	---YTVTADAADRIIKAINVAGTTPVGLNITQNTVGSIIKGG--NLLPVLNAGKSLTL	144		
Qy	376	KGGETQADKLTDDNNIGVVDNNTGL-----KVKLAKNL--SOLE	413		
Db	145	GNNAVA-----ANHFEDAPADNYTGLGNIALGGANAALIQSAAPSKITLAGNIDGGII	200		
Qy	414	TVSTRK-----NLTASEKVTYVSGNTEALQSGG-----LTFTPTTNASTDKIVYGTG	461		
Db	201	TVKTDAAALNGTIGNTNALATVNVGAG--TATL--GGAVIKATTTKLTNAASVLTL--	251		
Qy	462	LKFTDNSNT-----ALEDTT-----RITKDKIGFSNKAGTVDENKPYLDKD	502		
Db	252	-----TNANAVLTGAIDNTGDDNVGNLNGALSOVTDG-IGNTNSLATI-----	296		
Qy	503	KLKVGNSTLNNGLTVNNTIGGSKQIQVGADGKIFADVNV-----NVSNAAK-----F	551		
Db	297	--SVGAGTATLGGAVIKAT-----TTKLTDAASAVKFTNPVVVTVGAIDNTGNANNGIVTFT	350		
Qy	552	GTTRITEEIGFADAGKVKKSPYLDKKQLQVGGVKTIKDSGINAGDOKISNVKDATDD	611		
Db	351	CNSTVT--GNVGNTNALATVNVGAGL---QVGGVVK-----ANTINLTDN	392		
Qy	612	TDVVTYKOLKVOODADGALQSFSDIRDEKGQEFITSNLYSNGNTPNTFETITFAGEN---	668		
Db	393	ASAVTFTNPVV-----TGAID-----NTGNANNGI--VTFETGNSVT	428		
Qy	669	-GISISNDIAKGVKVGIDPENG-----LTPPKLVGSDKKOKTOLVIEQVASGNDTKNII	723		
Db	429	GDIGNTNALATVNVGAGTATLGGAVIKATTTKLTNA-----	465		
Qy	724	RGLSPTLPSITNAGVTRTEOGNTITSDEKSKAASIGDILNTGFNLKNNNSVGEVSTY	783		
Db	466	-----SVLTLTNANAVLTGAIDNTTGGD-----NVG-VLNLNGALSOVTCIGNTNSL	512		
Qy	784	NTVDFIDG-----NATTAKVT-----YDETNOTSK--VTVYDNNV	815		
Db	513	ATISVAGTATLGGAVIKATTTKLTDAASAVKFTNPVVVTVGAIDNTGNANNGIVTFTGN-	571		
Qy	816	DEKTIELTGDKGKTKNGVKTTLTTTNANGKATNFSTDDNALVNKADIAENLTLAKE	875		
Db	572	-----STVTGDIIGNTNSLATISVAGTATLGGAVIKATTT-----KLTNAASVLTLTNANA--	622		
Qy	876	IHTTKGTADTALQTFKVKKDGATDDETTIVGDKGTQNGKTVNTLKLKE-----NGLTVAT	931		
Db	623	-----VLTGAIDNTTGGDNGVGNLNGALSOVTDGIGNT	656		
Qy	932	NKDGTVFTGINTQSLKAGDSTTLNKGSLSKNPNASNEQIQVGADGVKFAK-----VD	984		
Db	657	NSLATISVG-----AGTATL--GGAVIK--ATTIKITNNAVSAVKFTNPVVVTVGAID	703		
Qy	985	K-GNSSTGI---DGFSTRITKQDQIGFTGANGSIDTTKPHLTKDKLKVGVEITNTGINAGG	1040		
Db	704	STGNANNGIVTFTGNSVTGID-IGNTNA-----LATVNVGAGTATLGGAVIKATT	752		
Qy	1041	KKITNIQSGDITQNSN-----DAVTGG-----RVYDLKTELESKINSAAKTAQNSLHEFS	1090		
Db	753	TKLTNAASVLTLTNANAVLTGAIDNTTGGDNGVGNLNGALSOVTDGIGNT--NSLATIS	810		
Qy	1091	VADEGNIHFTVSNPSSYDTSK---TSDVITTFAGENGITTKVNVGVVRVGDIDQTKGLTTP	1147		
Db	811	VG---AGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTG-----AVDNTTGT-----	855		
Qy	1148	KLTVGNNGNGIGVIDSQKQNTIT---GLSNTLANVTNDGAGHALSQGL---ANDTDKTR	1201		
Db	856	---GDNVG---VLNLNGALSOVTDGIGNTNSLATI--SVGAGTATLGGAVIKATTTKLTN	907		
Qy	1202	AASIGDVLNAGFNLOGNEAVDFVSTYDVTDFIDGNATTAKVTYDDTSKTSKVYDVYDND	1261		
Db	908	AASVLTLTNANAVLTG---AIDNTTGGDNGVGNLNGALSOVTDGIGNTNSLATISVGAG	964		
Qy	1262	NKTEVTSRDKLG---VKTTTLTKTSANGNATKFS-----AADGDALVKASDIA--THL	1310		
Db	965	TAT-----LGGAVIKATT--TKLTDAASAVKFTNPVVVTVGAIDNTGNANNGIVTFTGN	1015		
Qy	1311	NTLAGDITQAKGASOASSASYVDADGNKVYDSTDKKYQVNDKQQVDKKNKEVAKOKLV	1370		
Db	1016	STVTCNVGNTNALTATVNVGAGLLQVQ-----GVVKANTINLTNAS	1057		
Qy	1371	QAQOTPDGTLAQMNVKSVINKEQVNDANKKOGINEDNAFIKGLENAAKDTTKNAAVT--	1428		
Db	1058	AVTFT-----NPVVVTVGAIDNTG--NANNGIV-----TFTGNSVTGN	1093		
Qy	1429	VGDLNVAQOTPLTFAGDTGTTAKKLGETLTIKGG--QTDTKNLTDN-----	1472		
Db	1094	VGNTNALA-----TVNVGAGLLQVGGVVKANTINLTNASAVTFTNPVVVTVG	1141		
Qy	1473	-----NIGVVACTDGFVVKLAKDLTNLN---SVNAGGTTRIDEKGLSFSFDANGQAKAN	1521		
Db	1142	AIDNTGNANNGIVTFTGNSVT--TGDIGNTNALATVNVGA-----GIT--LQAGGSUAAN	1192		
Qy	1522	T-----PVLISANG--LDLGGKRISNI--GAAVDDNDA---VNEKQFNEVAKTVNNLNQOS	1569		
Db	1193	NIDFGARSTLEFNGPLDGGKAIPIYFKGATANGNALLNVTNLTTLTASHLTIGTV-AEI	1251		
Qy	1570	NSGASLPFV-----VTDANGKPIGTGPKQKAIKAGDAGKYHYHANANGVVPVOKGKPI	1622		
Db	1252	NIGAGNLFTIDASVGDVITLNAQIN-----FRARDSVLVLSNLTGTVGN---NIL	1299		
Qy	1623	TDADKLANLAHG-----KPLDAGHQVVASL-----CGN-----SDAI	1655		
Db	1300	LAADLVAPGADGTVFVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGV	1359		
Qy	1656	TLTIKSTLPIDTPTNGNANAGQAQSLPSLSAAQSSNAASVKDVLNVGNFLQTNHNOVD	1715		
Db	1360	NLEGIONVLI-----NKNADFTSTAPNAGAIQINDATVTTIDANN--NLNIPAGNIQ	1410		
Qy	1716	PVKAYDVFVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGVNGV	1775		
Db	1411	FAHADAQLVLSQNSGSDRITITLGANIDPDN-----DDEGIVIL-----	1448		
Qy	1776	LMPNGSLKAGKSADA--KTPTGLSLVNPNGKSGTSDAVALNNLSKAVFKSKDGTFTTT	1833		
Db	1449	---NSVTAGKKLTIAGGKTEGGAHKLOTILFKGA-GDC-----STAGTFTNT	1491		
Qy	1834	VSDGISIQGKDNSSITLSKDG--LNVGKGVISNVGKGTPTDAAANVOQLNEVRNLLGLGN	1892		
Db	1492	-----TNIVLDITGQLELG-----ATTANVVLFNDAVQLTGTGN	1525		
Qy	1893	AG---NDNAD-----GNQVNIADIKKDPNSSSSNRRTVIKAGTVLGGKGNNDTEKLA--	1941		
Db	1526	IGGFLEFNKAGMVTNLLNNVNVAGAVQ--NTGGTNGNGLI-----VLGASNLNRVNGIAML	1579		
Qy	1942	-TGGVQGVQVDKDGNA--GDLSNVNVVKTKQDGSKRALLATYNAAGQTNLTNPAEADRI	1999		
Db	1580	KVGAGNVTIAKGGKVIKGEIOGTGTNT-----LTLPAHFNLTG-----SINKT	1622		
Qy	2000	NEQIRFHVNDGNOEPVQVQNRNGIDSSASGKHSVAIGFOAKADGEAAVAIGROTQAGNQ	2059		
Db	1623	GGQALKLNFMMNGGVSQV-----GTAANSVGDITTAGAT	1657		
Qy	2060	SIATGQNAQATGDOSIATGTCNVVTGKHSGAIGDPTVKADNSYSVGNNGNFIQATQTDV	2119		
Db	1658	SFASVSNAGKATLGGTTSFANTFT--NTGAV-----TLAKGITSFAKN---VTAISF--	1706		
Qy	2120	FGVGNNTIVTESNVAIGLSNISAISAGTHAGTQAKKSDGTAGTGTTTTAGA-----TCTVKGF	2174		

Db 1707 --VANSATINFSNLAFSN--ITGG-----GTTLTGANOVTYGT-GSF 1747

QY 2175 AGOTAVGAVSGA--SGABRRIONVAAGEVSATSTDAVNSOLYKATOGIANATN 2227

Db 1748 TDTLTNTTFDGAAGSGNLIKSGSTLDSGVSTLAL-----VVTATN 1791

RESULT 2

IDB_YEBA_ECOLI STANDARD; PRT; 2003 AA.

AC P3366B; P76087; P76088; P76856; P76857; P76859;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.

GN YEBA OR B1401/B1405.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI_TaxID=562;

[1]

RN SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.;"

RL Science 277:1453-1474(1997).

[2]

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97251357; PubMed=9097039;

RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,

RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,

RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.,

"A 570-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 28 0-40.1 min region on the linkage map.;"

RT DNA Res. 3:363-377(1996).

[3]

RN SEQUENCE OF 464-2003 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=92190338; PubMed=1665988;

RA Moszer I., Glaser P., Danchin A.;

RT "Multiple IS insertion sequences near the replication terminus in

Escherichia coli K-12.;"

RL Biochimie 73:1361-1374(1991).

CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).

CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR

THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT

BETWEEN AMINO ACIDS 839 AND 840.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC -----

CC EMBL; AE000237; AAC74483.1; ALT SEQ.

DR EMBL; AE000237; AAC74487.1; ALT_SEQ.

DR EMBL; D90778; BAA15009.1; ALT_SEQ.

DR EMBL; D90778; BAA18880.1; ALT_SEQ.

DR EMBL; D90779; BAA18881.1; ALT_SEQ.

DR EMBL; X62680; -; NOT_ANNOTATED_CDS.

DR Ecogene; EG11307; ydba.

KW Hypothetical protein.

FT CONFLICT 489 489 I -> V (IN REF. 2).

FT CONFLICT 495 495 I -> V (IN REF. 2).

SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.8%; Score 556.5; DB 1; Length 2003;

Best Local Similarity 22.4%; Pred. No. 7.5e-13;

Matches 492; Conservative 243; Mismatches 776; Indels 685; Gaps 120;

QY 361 NCTGVKKKLGELTIKGGTQADKLTNNNIGVVTNNNTGLKVKL-----AKNLSGLE 413

Db 143 NNDVILDKTEKTLTIR---DSVFYTENADGTSLQDSNGRKATINLWQIDEANNITVALE 199

QY 414 TVSTKNLTASEKVTVCGSNGNTAELOSGGLTFTPTTNASTDKTGVYGTD--GLKFTDINSNTA 471

Db 200 GVS-----ADGATKWQYHNGELVITGDNATVNNNGKT--TYDGKDSGTSTEINGNGKV 251

QY 472 LED-----TTRITKDKIGFSNKGATVDENKP-----YLDKDKLKV---GNSLNN 513

Db 252 IQGDLDVSGGGHGIDITGDSATVDNK-GTWTVTDPESMGIQIDGDKAIVNNNGESTITN 310

QY 514 GG-----LTNNNT-----IGGSNKQ-----IQVGADGIGKFAVDN 542

Db 311 GGTGTQINGDDATANNGKTTVDGKDSGTGTGEINGNGKVIQDGDLDVSGGGHGIDITGDS 370

QY 543 VNVSNAAKFTGTRITE-BEIGFADADGKVDKSPYLDKKQLQVGG---VKITKDSGI-N 596

Db 371 ATVDNK---GTMVTDPEISG-----IQVDGQAVVNNNEGESAITN 408

QY 597 AGDKISNVKDATDDTDAVTYKOLQVQDADGALQSFSIRDEKGOEFTISNLSYNGNTP 656

Db 409 GGTGTQINGDDATANNGKT-----TVDG-----KDSGTGEIA---GNNGKVI 448

QY 657 NTFETITFAGENGISISNDIA---RGKVKVGIDPINGLTPTKLTGVSDDKGTQLVIEQ 712

Db 449 QDGDLDVSGGGHGIDITGDSATVDNKGTWV-----TDPE-SIGIQIDG-----DQ 493

QY 713 VASGNTKNIIRSLPTLSITNAGVVRTEQGNNTTSDDEKSKAASIGDILNTPMLKN 772

Db 494 AIVNNEGES-----TITN-GGTGTQINGDATANNNGKTTVDGKD--STGTKIAG 540

QY 773 NSNSVCFVSTYNTVDFIDGNATTAKVYDETNQTSKVTVYDVNVDEKTIETLTGNGKTN-- 830

Db 541 N---IGIVN-----LDGSLT-----VTGGHGVENIGDNGTVNNK 572

QY 831 ---KIGV---KTTTLTT---NANGKATNFSTTDNALVN-----A 862

Db 573 GDIVSDTSGIGVLINGEGATVNTGDVNVSNATGSIITNSKVSLSAGSMQVGFSTG 632

QY 863 KDAENLNTL---AKEIH-----TTKGTADTALQT--FKVKKDGATD-----DE 901

Db 633 VDLGNNSVTLAAKDKLVVQKATGINVSGDANTVNTGNVLVDKDKTADNAAEYFFDP 692

QY 902 TITVKGDTGQNGKTVNTLKLKGENGLTVATNKDGTVTFFGINTOSGL-----KA 949

Db 693 SVGINVYGSNNVTL-----DGKLTVSDSEVT-----SRQSLFDGSAKTSGLVVI 740

QY 950 GSTTLNKGQ---LSIKNP-ASNEQIQVGADGVKFAK---VDKGNSTGIDGTSRIKTD- 1001

Db 741 DGGNTVMNMGLELIGKKNALADGSOVTSRTGYSYTSVIVSVSGESSVYLNGLDITISGEF 800

QY 1002 QIGFTG-----ANGS-----LDTTPHUTKDKLKVGEVEITWTGIN-AGKK 1042

Db 801 PLGFAGVIRVDKALLEIGSGATLTMDIDISFEHHGTR-----TVEIQNLGFAFTVGEN 854

QY 1043 ITNIQSGDIT--ONSND-----AVTGRVYDLKTELESKINSAAKTAQSL-HEFS 1090

Db 855 TTGINSGTISLQNGKDPAPSPVILLATNGGSATNAGT-ITGKV-----TEQHSVFPNKS 908

QY 1091 VADEQGNHFTVSNPYSSYD-----TSKTSDDVTTFAGENGI-----TTKVNKG 1133

Db 909 TG--TSNSFFIENDVSSITGLVAQSNSTIINTDGGIIDLGRGSGVGMALADSTAEANOQK 966

QY 1134 VRVG---IDQTKGLTTPKLTIVGNNN-----GKGVIDSKDQNTITGLSNTLANVT--- 1181


```
Db 967 ITLDSMWVAND--TTAMRDIASNSAIDFGTGVGVGTDYSYSGAGKNATAINQLGGVITIY 1024
Qy 1182 NDGAGHALSOGLANDTKTRAASIGDVLNAG-ENLOGNGEAVDFVSTVDTDFIDGNATT 1240
Db 1025 NAGAGMA-AYGASN-----TVINQGTINLERNG-----NYD--DSLAANTLV 1063
Qy 1241 AKVTYDDTSKTSKVYVDVNDKNTIEVTSDDKGLGVKTTTLTKTSAN-GNATPKFSAADGDA 1299
Db 1064 GMAYVEHGTAINDOTGVINIVGTGQAFYNDGTG-----TIVNYGTICTFGVCQSGN 1115
Qy 1300 LVKASDIATHLNTLAGDIOTAKGASQASSASYVDADGNKVYDSTDKKYYQVNDKGOVD 1359
Db 1116 EYNTDDEFTSLIYTGDDTITRSGETVTLNKSAAV-----TDKLAGNVVNSGTLS 1164
Qy 1360 KNEVAKDKVAQAQTPDG-----TLAQMNVKSVINKEQVNDANKKOGINEDNAFIKG--LE 1414
Db 1165 GDQITVSSGLL--ENTSGGIINNLVKLDKGAVI-----KNAGVYTNVNDVSGGILN 1213
Qy 1415 NAAKDTK--TKNAAV-----TVGDLNVAQPTLTFAGDTGTTAKKLGELTILIKGGQDPT 1466
Db 1214 NAGEMTAQITNAGADSLVNTGTINKIVQAGVFNNSGVSVTGRMMS-----AGGV--F 1266
Qy 1467 NKLTDNNIGVAGTGDGFTVKLAKDLTNLNSVNAGGTRIDEKGISFVDANGQAKANTPVL 1526
Db 1267 NQQTGALMRGAALTGTAVANNEGTVNLGSSSEGN-----NTGMLE 1307
Qy 1527 ANGLDLGGKRISGAAVDDND--AVNEFKQFNEVAKTVN-NNLNQSNSGASLPFVVTVDAN 1583
Db 1308 VN----NNSAFNRRGEFLDNDKNAVHINQSGTLYNTGHMINSSSHNGA---VNMWGGN 1360
Qy 1584 GKPIN-GTIDGKPPQAI-----KGADCKYYHANANGV-PVDKDGKP-----ITDAD 1626
Db 1361 GRFINDETIDVSAKSLVYSANNAGDQNAFFWNQDNGVINFDHDSASAVKVFHSFIAQND 1420
Qy 1627 KLANLAHKGKPLDAGHGVVVASIGNSDAITLTIKSTLPQIDTPTNTGNANAGQASLP 1686
Db 1421 GIMNISGTG-----AVAMEGDKNA-----QLVNNGTINLG----- 1450
Qy 1687 SAAQSNAAKSKVQVNLVNGFLOTNHNQVDFVKAYDTVN-FVNGTGADITSVRSADGTMSN 1745
Db 1451 -----TAGTTDTGIMGMOLDANATADAVIENNGTINIFANDSFA-----FSVLGTGVH 1498
Qy 1746 ITVN-TALAATDDGDNVLKAKD-----GKFYKADDLMPN----- 1779
Db 1499 VVNGTIVYIADGVGTSGLIKQGDGINVEGMNNGNSSEVHYGDTLPLDPVKPNTSVTS 1558
Qy 1780 GSKAGKASDAKPTGLSLVNPAGKSGTGDA----VALNLSKAVFKSKDGTITTTVS 1835
Db 1559 GSDAGGSMNLNGYVVGTVNNGSAGKLKVNNASMNGVEINTGFTA-----GTADTTVS 1612
Qy 1836 SDGISIOGKD-----NSSITLSKDLNKGKVISNKGKGTKDTDAANVOQLNEVRN 1886
Db 1613 FDNV-VEGSNLTDAADATSTSVVWTAKGSTDASGNDVDVMSKNAY-TDVATDASVNDIAK 1670
Qy 1887 LLGLGNAGND-----NADGNQVNIADIKKDPNCGSSSNRTVTKAGTVLGGK 1933
Db 1671 ALDAGYTNNELFTSLNVGTTAEALSALKQV-----SGSQAT-TVFREARVL----- 1715
Qy 1934 NNDTEKLTATGCVQGVQVQDKGDNAGDLSNVVVKTKQKSKKALLATYNAAGOTNLTNPA 1993
Db 1716 SNRFSMLADAAPKVG-----NGLAENNVAK-----GDPAELG-----NN-- 1750
Qy 1994 EADRINEQGRFFHVNDGNGQEPVVOGRNGIDSSASGKHSVAIGFOKADGEAAVAIGRQ 2053
Db 1751 -----TEYDMLALRKITIDLESQTMISLEYGI-ARLD----- 1780
Qy 2054 TQAGNOSTAIGDNAQATQDQSIAGTGNVYTGKSHGALGDPSTVKADNSYSGVNNQFID 2113
Db 1781 -----GDGAQKAGD-----NGVTGYSQFGLKHQMSQGMNWN----- 1816
Qy 2114 ATQTDVFCVGNHNTVTRBSNVALGNSAISAGTHAGTQ-----AKKSDGTAGTTT- 2164
```

```
Db 1817 ALRYDVNHLDS-----SRSTAFG-NTNKTADTVDVKQOYLEFRSEGAKTFFPSBGLKVTP 1869.
Qy 2165 -AGA--TGTVKGFAGQAVGAVSVGA-SGAERRIQNVA-----AGE-----VSATSTDAV 2210
Db 1870 YAGVKLRHTLEGGYQERNAGDFNLNMNSGSETAVDSIVGLKLDYAGKDGWSASATLEGGP 1929
Qy 2211 NGSOLYKATQCIANATWELDHRIHONENKANAGISS 2246
Db 1930 NLS--YAKSQRTASLAGAGSQHFNVDDQKGGGINS 1963

RESULT 3
OMPB_RICJA
ID OMPB_RICJA STANDARD; PRT; 1656 AA.
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
```

Query Match 4.2%; Score 492.5; DB 1; Length 1656;
Best Local Similarity 22.7%; Pred No 1.1e-10;
Matches 419; Conservative 198; Mismatches 690; Indels 539; Gaps 93;

```
Qy 656 PNTFETITFAGENGISINDIAKGVKVGIDPPLINGLTTPKLTGSKDKGKTLVIEQVAS 715
Db 68 PNA--VITANANNGINLNTPAGS-----FNGFLF-----SNANLAVTVSEDTTL 110
Qy 716 G--NDFKNIRGLSPTLPSTNACGVRT-TEQGNITVTSDEKSKAAASIGDILNTGFNLKN 772
Db 111 GFINNAANANRFLTLT---DAGKTLTITGGTITNVQSAATHNAQIVAKFNGGAIAAN 166
Qy 773 NSNSVGFSVTNTVDFIDGNATTAKVTYDETNOTSKVTYDVNDVEKTLTGDKNGKTNK- 831
```


Db 167 ND-----LSGLGTIDP---GAAASTLVFLDANPTTQKAPLILADNALI-VNGANGTLNVT 217
QY 832 ---IGVKTTLTTTNA-----NGKATNFSTDDNDALYNKADIAENLNTLAKEIHITKGTAD 884
Db 218 NGFIOVDSKFATVKAINIGDGGEGMENTNATNA--NALNLQAGGTTI--NFNGTDTGTR 273
QY 885 TALOTFKVKKGATDDETITVKGDFQNG-KIVNTLKLKGBGLTVATWKNKGTVTFPGINT 943
Db 274 LVL-----LSKNGAATDFNYTSGLGNLKIIEIENVAI---NGOLIANAGPANAVIGTNN 326
QY 944 QSLKAGDSTTLNKGSLKSNPASNEQIQVGADGVKFAKVDKGNSTGIDGTSRIKDOI 1003
Db 327 GAGRAAG-----FVSVNDNGKAAT-IDGQVYAKDMVI 357
QY 1004 GFTGANGSLDTTKPHLTDKLKVGEIETNTGINAGKKITNIQSGDITQNSDAVTTGR 1063
Db 358 QSANANG-----QVNERHIVDVGIDGTAFKTAASIVAITQNSNEFTT--- 400
QY 1064 VYDLKTELESKINSAAKTAQNSLHEFSVADEQ---GNHFT--VSNPYSSYDTSKTSIDVIT 1118
Db 401 -----DFGNLAAQVTPDPTMTLTGN-FTGDANPN-----GNTAGVIT 436
QY 1119 FAGE-----NGITTKVNGKVVRVGDIDTKGLTTPKLTGVGN--GRK 1158
Db 437 FRANGTLASADANVANVTNNITAI PASG---VGVVOLSHTTAEDLUGNAGSVFKLAG 493
QY 1159 IVDSK-----DQONTITGLSNTLANVTNDGAGHAL-SOGLANDTDK 1199
Db 494 TVINGRVNQTVLVGGVLAAGAITLDSATITG-----DIGNGGGGAALQSITLANDATK 547
QY 1200 TRAASGCDVLNAG---FNLQNGEAVDFSTVD--TVDFIDGNATTAKYTYDDTSKTSKV 1254
Db 548 TLTGLGANITSANGGTFINFGANGGTTIKLTSTQNNIYVDCDLAIATDQTGVVDASSLT--- 604
QY 1255 VYDVNVNDKTIETVSDKLGKVTTL-----TKTSANGNATKFSAADGDALVAKASDIA 1307
Db 605 ----NAQTLTISCTIG-IICANNFTLQGNIGSSKTTLAG-----GNVAINELVIG 650
QY 1308 THLNTLAGDIQTAKGASQASSASVYDADGNKVIYDSTDKKYQVNDKQGVNDKNEKAVKD 1367
Db 651 NN-----GSVQFAHNTYLITRTTN--AAGGKIIFNPV-----VNNNTTLAAG 691
QY 1368 KLVAAQATPDGTLAQNNVSVINKEQVND--NKKOGINEDNAF IKGLENAAKDTKTNA 1425
Db 692 TNLGSAANP---LAEINFGS---KGARADTVLNVGEGVN-----LYATNITTTDANVGSF 740
QY 1426 AVTVGDLNAVAOTP-----LTFAGDTGTTAKKLG-----ETLTIKGGOT 1464
Db 741 VFNAGKNIVSGTVGCGQGNKFNWALDNGTTVKFLGNATFNNTTIAANSTLQISGNVT 800
QY 1465 DTNKLTDNNIGVV--AGTDGFTVYKAKDLTNLSV-----NAGGPRIDEKGISFVDANG 1516
Db 801 ADPIASADGTGIVEFVNTGPI NVTLNKQAVPVNALQITVSGPGNVVVEIG---NAGN 856
QY 1517 QAKANTPVLISANGLDLGGKRISNIGRAVDNDNAVNFQFNEVAKTV-NNLNMQNSGASL 1575
Db 857 YIGAMTDITAFENSSIGA--VLFLPSGIPFNDA-----GNTIPLTIKSTVGNETAEGFSV 909
QY 1576 PFVVTDANGKPIGTGPKPKAIKAGDGKYYHANANGVPVDKGPITDADKLANLA--- 1632
Db 910 PSVI-----VSCVDS-----VIADGQVIGQONNIVGLGLGS 940
QY 1633 -----AHGKPLDAGHOVVASLGGNSDAITLTNKTSLPQIDTPNTGNANAGQAQSLPSLS 1687
Db 941 DNGIIVNATTLVAG---IGTINNOCVTILSG-----GVNTPGTIVYGLG---TGIG 986
QY 1688 AAQOSNAASVKDVLNVLNQLNHNQVDFKA-----YD---TVNFVNGTG 1730
Db 987 ASKFQVOTFTDYNLNLGNIATNTINDGVTVTTGGIAAGGIAGTDFDKITILGVSNGN- 1045
QY 1731 ADITTSVRSADGTMSNTVTALAAATDDCGNVLKAKDGKF-YKADLMLPNGLSKAGKSAS 1789
Db 1046 ---ANVRFADGIFSN---STSMIVT-----TKANNGTVTYILGNAFVGN-----IG 1084

QY 1790 DAKTP-TGLSLVNPAGKSGTGDA-----VALNLSKAVFKSKDGT 1830
Db 1085 DSDTPVASVRFTSGNNGAGLKGNIQVDFGTGYNLGINSVNLGSGTAINKIDLLT 1144
QY 1831 TTTVSDGSIQCKDNSSI---TLSKDLGNVGKVI---SNWKGTKDTDAAVYQQLNEVR 1885
Db 1145 NLTUFAAGTSTWG-NNTSIETTLTLANGNIGHIVIAEGAQVNAATTTGTTTINVQ----- 1197
QY 1886 NLLGLGNAGDNADGNQVNIADIKKDPNCGSSNRVIRKAGTVLGGK--GNNDTEKATG 1943
Db 1198 -----DNANA-----NFSGTQTYILIOGGAAREFNGTLGGPNTVTGTSNR 1235
QY 1944 GVOVGVDKDCNANGDLSNVVKTKQDKSKKALLATYNAAGQTNLTNNPAEADIRNEQG 2003
Db 1236 FVNYGLIRAANQD-----YVITRTNNAENIV-----TNDITNSPPGGAPGVQNV 1280
QY 2004 IRFFH-VNDGNQBPVVQGRNGIDSSAGSKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIA 2062
Db 1281 TTFVNATNTAAAYNNLLAKNSADSANFVGTIVTDTSAAITNAQLDVAKOIQALGNRLGA 1340
QY 2063 IGDNAQTGDSIAIGTGNVVTGKHSAGIDPSTVKA-----DN-SYSGVNNNQFIDATQ 2116
Db 1341 LR-----YLGTPEMV-GSEAGAI--PAAVAAGDEAVDNVAVGIMAKFPYTDHQ 1386
QY 2117 TDVFGVGNNTVTESVALGNSAISAGTHAGTQAKKSDGTAGTTTGTAGATGTVKFGAG 2176
Db 1387 SKKGGI-----AGYKAKTTGIVIGLDTLAN-----N 1412
QY 2177 QTAAGVAVSGASGAERRIIONVAAGEVSATSDAVNG--SOLYKATOGIAN--ATNELDHR 2232
Db 1413 NLMIGA-AIGITTKTDIKHODYKKGD---KTD-VNGFSFSLYGAQQFVFNFAQGSATFS 1466
QY 2233 IHNENK-----ANAGISSAMAMASMPQAYIPGSRMTVGGIAT---HNGOGAVAV-- 2279
Db 1467 LNOVKNSQRYFPDANGNKSQITAGNYD-----NMTFGGNLTVGYDYNAMQGVLVTP 1519
QY 2280 --GLSKLSDNGQWVFYKING-----SADTQGHVGAAYGAG 2311
Db 1520 MAGLSYLKSSDE-NYKETGTTVANKQVNSKFSRDTDLIVGAKVAGG 1564
RESULT 4
HLVA_PROMI
ID HLVA_PROMI STANDARD; PRT; 1577 AA.
AC P16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HEMOLYSIN PRECURSOR.
GN HPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxId=584;
[11]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC STRAIN-ISOLATE 477-12;
RX MEDLINE=90170827; PubMed=2407716;
RA Uphoff T.S., Welch R.A.;
RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin genes (hpmA and hpmB) reveals sequence similarity with the Serratia marcescens hemolysin genes (shla and shlb).";
RL J. Bacteriol. 172:1206-1216(1990).
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA REQUIRES HPMB FUNCTION.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA MAY BE RESPONSIBLE FOR PORE FORMATION.

CC -!- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M30186; AAA25657.1; -;
DR PIR; A35140; -;
KW Hemolysis; Toxin; Outer membrane; Signal.
FT SIGNAL 1 29
FT CHAIN 30 1577 HEMOLYSIN.
SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

Query Match 3.9%; Score 450.5; DB 1; Length 1577;
Best Local Similarity 21.1%; Pred. No. 3.3e-09;
Matches 389; Conservative 249; Mismatches 605; Indels 603; Gaps 95;

Qy 593 SCINAGQKISNVRDADTDDAVIYKQKQVQDADGALQSFIRDEKQGEFTISNLYSN 652
Db 44 SAVNGGTQVINIV---TPNNEGISHNQYQDFNVGKPGAV--FNNALEAGQSQLAGHLNAN 98
Qy 653 GNTPEFTETIFAGENGISINDIAKGVKVGIDPINGLITPKLTGVGSDKDKTQLVIEQ 712
Db 99 SN-----LNQQA-----ASLIINE 112
Qy 713 VASGNDT-----KNIIRGLSPPLSITNAGVTRTEQGNITTSDEKSKAASIGDIL----- 764
Db 113 VVSRNPSPLLQQQEVFGIAAEY-VLSNPNGITCDGCGIFNS-----RSSLVVGNPLFENG 167
Qy 765 -NTGFNLKNSN--SVGPVSYNVDFD-----GNATTAKVYDETNQTSKYTYD 812
Db 168 QLKGYSTLNNLNLSLG-KNGLNTTGLLDLAPRIDSRGKITAAEISAFTEQNTFSQHFD 226
Qy 813 VNVDEKTIELGDGKTNKIGVKTITTTTNNANG-KATNFSTTDNDALVNAKDAENLNT 871
Db 227 ILSQKPVSAIDSYFFGSMQSGRIINTAEGSGVKLAGKFTADNDLSVK-----ADNIQT 282
Qy 872 LAKEHTTKGTADTALQTFKVKDGATDDEY-----ITVGKDGTONQTVNLTCLKGENGL 927
Db 283 -----DSQVRVDSYDKGSENYQNYRGGITVNNSG--SSQPLTKTELKGN-- 326
Qy 928 TVATNKDGTVFGINTQSLKA-----GDSITLTKDGLSIRKPASNEQIQVGADGVKFAKV 983
Db 327 -----ITLVASSHNQIRASDLMGDDITLOGADLTI-----DGKQLQ----- 362
Qy 984 DKGNSTGIDG-----TSRIT--KQIQGYGA-----NGSLDTTKPHLTCKLKVGEV 1029
Db 363 ---QKETIDIRWFYSWKYDYVTEKEQIQQIGSQIDAKNNATLTATKGDVTLDAK----- 415
Qy 1030 EITNTGINAGKKITNIQSGDITONSNDVATGGRVYDLTKLESKINSAAKTAQNSLHEF 1089
Db 416 -----INAGNNLAIN-----ANKDTHINGLVEKESRSENGKRNHTSLESGW--- 459
Qy 1090 SVADQGGNHFTVSNPYSDYTSKTSQDITTFAGENGITTKVAKGVVVRVGDQTKGLTT--P 1147
Db 460 -----SNHQY-ETLKASELT--AGKD-----LGLDAQGSITAQGA 492
Qy 1148 KLTGVNNNGKIVIDSKQONTITGLSNTLANVYNDGAGHALSOGLANDTDK----- 1199
Db 493 KLIHANE-----VLNNAKNNIINLVQKTKNDKTVTDN-----HYMGGIGGQNNNNNQOV 545
Qy 1200 ---TRAASIGDVLNAGFNLOGNGEAVDFVSTVDYDFIDGNATTAKVYDDTSKTSKVYV 1256
Db 546 SHATQLTADGQLLLAADN-----NVTGSOVKGNQGNQAFVKTTQ 585
Qy 1257 DVNVDNKTIETVSKKLGKVT---LPKTSANGNATKFSAADGDALVKASDIATHNL 1313
Db 586 DVVIDNALSETIS--KIDERTGTAFNITKSSHKNETN----- 621

Qy 1314 AGDIQTAKGASQASSASYVDADGNKVIYDSTDKKYYQVNDKQVNDKNEKAVAKDLVAQA 1373
Db 622 ---QSTG-SELISDAQLTVVSGN-----DNNVIGSLIK-----SADKLGH- 659
Qy 1374 QTPDGTIAQNNVSVKINKEQVNDANKQGIN-----EDNAFIKGL-----ENAKADT 1420
Db 660 ---SLGDINVKSAQVTKIKDEKTSIAITGHAKVEDEKQYSAGFHTHTNTKSTST 714
Qy 1421 KTKNAAVTVDGLNAVAQTPLEFAG-DTGTTAKKLGETLTIKGG-----QDTNKLKTNNI 1474
Db 715 EQANSTISGANVDLOANKDVTTFAGSLDKTTA---GNASITGDNVAVSTENKKTOTDND 770
Qy 1475 GVWAGTDGFTVKLAKDLTLNLSNVNAGTRID---EKGISFVDANGQAKANPPV-----LS 1526
Db 771 TTISGGFSYT-----GGVDKVGSKADFQYDKQHTQTEVTKNRGSQTEVAGDLTIT 820
Qy 1527 ANGLDL-----GGKRISNIGAADVDDNAVNEKQFNEVAKTVNNLNQNS 1571
Db 821 AN-KDLLHEGASHHVEGRYQESGENIQHL--AVND-----SETSKT-DSLNVGIDV 867
Qy 1572 GASLPFVVTDANG--KPINGTDGPKQAIKADGKYYHANANGVPVDKDG-----KPI 1623
Db 868 GVNL-----DYSGYTKPV-----KKAIE-----DGYNTTKPGNNTDLTKV 904
Qy 1624 DADKLANLA--AHGKPLDAGHQVVASLGGNSDAITLNIKSTLPDPTPTGNANAGQA 1681
Db 905 ARDAIANLANLSLETNPNVGEVIGKGGSQSQSDQAVST-----SINAKGIDIDSNN 959
Qy 1682 SLPSLSRAAQSNAAKSVKDLNVGNENLQTNHNQVDFVKAYDTVNFVNG-----TCAD 1732
Db 960 KLHD-----QGHYOSTQEGISLTANTHTSEATLD--HQHTTTHETKGGQIGVSTK 1013
Qy 1733 ITSRSADGTMSNTVTNTALAAATDDGNNLIIKAKDGKPYKADDLMPNGLSKAGKASDAK 1792
Db 1014 ITVAIKGEGQ-----TTDNALMET-----KAGSOFTSNGDISINVENAHYEGAQFD 1061
Qy 1793 TPTGLSLVNPAG-----KGSTGDAVALNLSK-----AVFKSK 1826
Db 1062 AQKGTVINAGGDLTLAQATDTHSESQSNVNGSANLKVGTTPESKDYGGFAGTTTHSK 1121
Qy 1827 DGTIT---TTVSSDGI-----SIQKDNSS---ITL-----SKDGLN 1857
Db 1122 EOTAKVGTITSGOIELNAGHNLTLOGTHLSSEODIALNATNKVDLOSASSEHTEKGN 1181
Qy 1858 VGGKVISNVGKTKDXTDAANVOQLNEVNNLLG-----LGNAGNDNADGN--- 1901
Db 1182 LSGVQAGFGKKMTD-DASSVNGLSGAQFAIGKQDEKSVSREGGTINNSGNLTNGSVH 1240
Qy 1902 ---QVNLADIKDPNSG-----SSSNRTVKA--GTVLG--GKGNDDT-----EKLAT- 1942
Db 1241 LQGAQVNSKDTQLTSQSGDIEITSAQSTDYAKNNMCTDIGFMGKKTNTNPPKEVTEKPATS 1300
Qy 1943 ---GGVQGVGD-----KQGNANGDLSNVVVKTKD-----GSKKALLATYNAAGOTNY 1987
Db 1301 IHNIGKLLVNVEDQKTSHQATLETGTLTINSKDLTLSGANVTADSVTGNVGGSLNI 1360
Qy 1988 LTNPAEAIIDRINQGIIR--FFHVNDGNQEPVVOGRNIDSSASG-----KHVSATGF 2038
Db 1361 ASQKES---DRHVTGVNVGYNHTNDPKSSQVKN-----TAKAGGSLLKTIKOTIDSGI 1412
Qy 2039 QAKADGEAAVAIGRQTOAGNOSIA---IGDNAQATGDOSIA-IGTG--NVVTGKHSAGI 2091
Db 1413 KSTSD-----AISDKYNSLSSTIAADKTCISDETAKIDQGFCKVGNGLKNIIVTC----- 1461
Qy 2092 GDPSTVKADNSYSVGNNNQFIDATQDVFVGNNITVTESVALGNSAISAGTHAGTQ 2151
Db 1462 AEGHTANAD-----IKVTHVDNDVATKTTSLTSLNNDLSLVN----- 1498
Qy 2152 AKSDGTAGTTTTACATGTVKGFAGQATVAGVAVGASCAEBRIQNV 2197
Db 1499 -----GSTKLTA---EIVSQQGVLDGGSSV-----KLENI 1527

Db 1558 VLDKXHDLTVTASQADNRSLK---AGHDFTVQAOQIDNSGTMAAGHADTLKAP-HLRN 1613
Qy 1218 NGEAV-----DFVSTYDTVDIDGNATTAKVTYDD 1247
Db 1614 TGQVAGHDHIIINSKLENTGRVDARDIALDVAADFTSGSLAEHDAITLTAQGTORD 1673
Qy 1248 -----TSKTSKVYVDV--NVDNK-TIETVTSKKLGVK 1276
Db 1674 LVVDODHILPVAEGTLRVKAKSLTTEIETGNPGLIAEVEQENIDNKQAIYVGKD----- 1727
Qy 1277 TTTLTKTSGANGNATKFSADGDALVKASDIATHLNTLAGDITOTAKGASQASSSAYVDAD 1336
Db 1728 ---LFLSSAHNV---ANEANALLWA-----AGELT-TVKAQNIITNKRAALIEAG 1769
Qy 1337 GNKVIYDSTDKKYYQVNDKGQVKNKEVAKD-KLVAQAQTDPDGTIAQMNKSVINKEQVN 1395
Db 1770 GNARLTAIV---ALLNKLGRIRAGEDMILDAPRIENTAKLSGEVQKGVQDVGGGEH-- 1823
Qy 1396 DANKQGINEDNAFTKGLUENAKDKTKTNAAVTVGDILNAVATPLTFAGD--TGTAKK- 1452
Db 1824 --GRWSGICYVNYWLRA--GNGKKAGTTAAPWYGDLTA-EQSLIEVGKDLVYNAGARKD 1878
Qy 1453 ---LGTFLTKGGOTDNKLTNN--IGVVAGTGDGTVKLAKDLTNLNSVNAAGTTR-- 1503
Db 1879 EHRHLNEGVIQAGGHGHIIGGVDDNRSVVRIVSAMIEKTPLPVSLTALD--NRAGLSA 1936
Qy 1504 -----IDKRGISFV-----DANGQAK---ANTPVLNAN 1528
Db 1937 TWFQSTVELLDYLLDONRYEYIWLGYPTYTSEWSYNTLKNLDLGVQAKPAPTAPMPKAP 1996
Qy 1529 GLDLCGKRISNIGAANDNDVAFKQFNEVAK-----VPVDDKDGKPIITDADK----- 1627
Db 1997 ELDLRGHTLES-----AEGRKIFGEYKKGLEYEKAKMAVQAVEAYGEATRRVHDQ 2047
Qy 1569 SNSGASLPFVWTANGKPIG-----TDGKPO----- 1595
Db 2048 LCQRYGKALGGMDAETKEVDGIIQFEADLRTVYAKAQDQATIDAETDKVAQRYKSQIDA 2107
Qy 1596 -----KAIGADGKYHYANANG-----VPVDDKDGKPIITDADK----- 1627
Db 2108 VRLQAIQPGVTLAKALSAAALGADWRALGHSQLMQWRKDFRAGKRGAEITAFVPEQTVLA 2167
Qy 1628 -----LANLAH-----GKP--LDAGHOVVASLGGNSDAITLTIKSLTP----- 1665
Db 2168 AGAGLTLSNGATHNGENAAQNGRPEGLKIGAHATSVSFGFDALRDVGLKRLDIDDAL 2227
Qy 1666 -----LQDTPNTGNMA----- 1677
Db 2228 AAVLVNPHITRIGAAQTSADGAAGPALAQARQAPETDGVNDARGLSADALASLASL 2287
Qy 1678 GOAQL-----PSLSAAQOSNAASV-----KDLVNVGFN--- 1706
Db 2288 DAAQGLEVSGRNQAQVADAGLAGPSAAVPAVGAADVGPVGTQDVQDPVAVVLEQPV 2347
Qy 1707 -----LQTNHNVQDFKAYDIPNVFNGTG-----ADITSV 1736
Db 2348 ATVRVAPPAVALPRLFTFRKIDQSKFYGSRYFEQIGYKPDRAARVAGDNYEDTTLV 2407
Qy 1737 -----RSADGTMSTNV--NTALAA--TDDGNNV-----LIRAKDGKF-YK 1772
Db 2408 REQVRALGGYESRPLVRGVVALVKALMSAGTVGKALGLKGVGVAFTAQQLKQADRFVWY 2467
Qy 1773 ADDLPNPSLKGAKSADAKPTTGLSLVNPNAKG-----STGDV-----ALNKL- 1818
Db 2468 VDTVIDGQKVLAPRLYLTATREQGIT--DQYAGGALLIASGSDVTVNTDGHVDVSVNGLI 2525
Qy 1819 -----SKAVFKSKDGTTTTTTSVSDGISIGOKD-----NSSITLSKDXGLNV 1858
Db 2526 QGRSVKVDAGKGVVAVDSKAGGAGIEADDEVDVSGRDIGEGGKLRGKDVPLKADTKVY 2585
Qy 1859 G-----GKVISNVGKGTDKTDAAVQ-----QLNEVRNLLGL 1890
Db 2586 ATSMRYDDKGRUAAR-GDCALDAQGQLHIEAKRLTAGATLKGKVKVLDVDDVK-LGGV 2643

Qy 1891 GNAGNDNAD-----GNQVNIADIKKDPNSGSSNR--TVTKAGTVLGGKNNNDTEKIAT 1942
Db 2644 YEAGSSYENKSSSTPLGSLFAILSTTETNQSAHANHYGTREAGTLEGKMQNLETE----- 2699
Qy 1943 GG-----VQGVYDKDGNANGDLSNVVVKTKD-----GSKKAL 1975
Db 2700 GGSVDAHAHTDLSVARDARFKAADFAHAHEKDKVRQLSLGAKVGAGGYEAGFSLGSEGL 2759
Qy 1976 LA-----TYNAAQGTNYLTNNPAEIDRINEQGIREFHVNDGQEPVQVQGRNGIDSSAS 2029
Db 2760 EAHACRGMTACAEVKVGYRASHEQSS-----ETEKSYRNANLNFPGGS--VEAGNVLDIGGA 2814
Qy 2030 GKHSVAIGFOAKADG-----EAAVAIGRQT-QAGNOSIATGDNAQA----- 2069
Db 2815 DINRNYGGAAGNAGTBEALMRRAKKVSTKYVSEQTSQSSGWSVEVASTASARSLIT 2874
Qy 2070 ---TGDSIA--TGTGNVTVTK-----HSGAIGDPSTVKAD---NSY 2103
Db 2875 AATRLGD-SVAONVEDGHEIRGELMAAQVAABATQVLTADTAVALSAGISADFDSSHRSR 2933
Qy 2104 SVGNNOF-----IDATQTDV-----FGVNNITVTESNVAL-----GSN 2139
Db 2934 STSQNTQVLGNLSTIATEGDTLVGAKFGGQDQVSLKAAKSVNLMAAESTFESYSESHN 2993
Qy 2140 SAISAGTHAGTQAKSDGCTAGCTTTTAGATGTVKGFAGTAVGAVSVGASGAERRIQNVAA 2199
Db 2994 PHASADANLGNANVQGVGLGTAGMGTSHQITNETGTYAG-TSVDAAVY-----SIDA 3047
Qy 2200 GFVSATSDAVNGSOLYKATOGIANATNDELDRHITHONENKANAGISSAMAMASMPQAVIP 2259
Db 3048 GKDLNLSGRVRGKHVLDVEGDINATSKQDER---NTNSSGGGWDASAGVA-----IQ 3098
Qy 2260 GRSMV-----TGGIATH-----NGQGAVALGSLKLS 2285
Db 3099 NRTLVPVGSAGFNTBHDNSRLTNDCAAGVVASDGLTGHVKGANLTGTATADLS--- 3155
Qy 2286 DNGQWFKINGSADTQ 2301
Db 3156 --GRGNLKVQDGAVNAQ 3169

RESULT 6

OMP_B_RICPR
ID OMPB_RICPR STANDARD; PRT: 1643 AA.
AC Q53020; Q9ZCM0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiellae; Rickettsia.
OX NCBI_TaxID=782;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=BRINL;
RC MEDLINE=91045972; PubMed=2122457;
RX Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
surface-layer protein of Rickettsia prowazekii: presence of a
truncated identical homolog in Rickettsia typhi."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=BRINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompB of Rickettsia prowazekii."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

[3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E;
 RA MEDLINE=99039499; PubMed=9823893;
 RX Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN=REINL;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 [5]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RC MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent mutant
 RT deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M37647; AAA26390.1; ALT_INIT.
 DR EMBL; AF161079; RAD42234.1; -;
 DR EMBL; AJ235273; CAAL5140.1; -;
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1328
 FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN REINL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN REINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN REINL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTQEAPLTGA -> INSRSSSVHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 3.8%; Score 445.5; DB 1; Length 1643;
 Best Local Similarity 21.4%; Pred. No. 5.1e-09;
 Matches 392; Conservative 216; Mismatches 713; Indels 509; Gaps 88;
 QY 776 SVGFV--SYVTNTVDFIDGNATKATVYDETQNTSKVTYD-VNVDEKTIETLTGDNKTNKI 832
 DB 13 SAGLVTAATATVAGSGVAMGRAMQYKRTTNAATTFDGIQFD-----QAAGAN-----I 63
 QY 833 GVKTTTLTTTNGAKATNFSTTDNDA-----LVNAKDAIENLNTLAKEIHTTKGTADTALQ 888
 DB 64 PVAPNSVITANANPIT-FNTPNGHLNSLFLDTANDLAVTINE-----DTTLGFTINIAQ 117
 QY 889 TFK-----VKKDGTADDE---TITVKGDTGQ--NGKTVNTLKLKGENGLTVA 930

Db 118 QAKFFNFVAAGKILNITQGGITVQEAASNTINAQNALTQVHGAAINANDLSGLSTIFA 177
 QY 931 -----TNKDGTVTFGIN-----TOSGLKAGDST-----TLNKD- 958
 Db 178 AAPSVEFNILNITQEAPLTLGANSKIVNGGTLNITNGFTQVSDNTTFAGIKTINIDD 237
 QY 959 --GL---SINKPASNEQIQVAGDGVKFAKVDKGNSSSTGIDGTSRI-----TKDQIGFT 1006
 Db 238 CQGLMFNSTPDAANTLNQVGGNTINF-----NGIDGTCKLVLSKNGAATEFNVT 288
 QY 1007 GANGSLDTTKPHLTOKLKVGEVEITNGI-----NAGGKKITHQS 1048
 Db 289 GTLGG-----NLK-GIIELTAAVAGKLISOGGAANAVIGTDNGAGRAAGFVS 336
 QY 1049 GDITQNSNDVAVTGRVYVDLKTLESKINSAAKTAQNSLH-----EFSVADE-----QG 1096
 Db 337 VD--NGNAATISQVYAKNMVQSANAGQVTFEHIYDVGLGTTNFKTADSKVIITEN 393
 QY 1097 NHFTVSNPYSSYDT-----SKTSDVITFAGENGI---TTKVNGK 1132
 Db 394 SNFGSTN-FGNLDTQIVVPDPTKILKGNFIGDVKNNGNTAGVITFNANGALVSASTDPNIA 452
 QY 1133 VVRV-----GIDQTKGLTTPKLFVGNNN-----GKGVIDSKQGNITGLSNTLA- 1178
 Db 453 VTNINAIEAGAGVWELSGHIAELRLGNGSGIFKLADGTVINGPVPNQNALMN-NNALAA 511
 QY 1179 -----NVTNDGAGHALSQ-GLANDTDKTPRAAS-----IGDLNAGNLOQNGE 1220
 Db 512 GSTOLDGSAITDIDGNGVGNALOHITLANDASKILALDGANIGANVGGAHFQANGG 571
 QY 1221 AVDFVSTYDFVDFIDGNATKATVYDDTSKTSKVYVDVNDKTIETVTSKDKLGKVTTL 1280
 Db 572 TIKLTNTQNNI-----VVFDDLTITDKTGVDASSL 603
 QY 1281 TKTSANGNATKFAADGDALVKASDIATHLNTLAGDTQAKGASQSSSASYVDADGNKV 1340
 Db 604 T-----NNQTLINGSIGTVVANTKTLLA-QLN--IGSSKTLNAGDVAINELVNIENGSVQ 656
 QY 1341 IYDST--DKKYVQVNDKQVDKNEKVAKDLVQAQPPDCTLAQMNVKSVINK-EQVND 1397
 Db 657 LNNTVLTITKINAAGQI---IVNADPLNTNTLADGT-----NLGSAENPLSTIIFA 708
 QY 1398 NKKOGINEDNAFIKLENAAKDTTKTNAAVTVGDLNNAVAQTPLTFAG----- 1444
 Db 709 TRAAANADSLNVGKGNLVANNITND--NVGSLHFRSGGTSIVSGVGGQGHKLNNL 766
 QY 1445 --DTGTTAKKLGETLTIKGG-----QTDINKLTD-----NNIGVV--AGTDGFTVK 1486
 Db 767 ILDNGTTVKFLGDT-TFNGGKTIEGKSILQISNNYTTDRHVESADNTGTLEFVNTDPTVT 825
 QY 1487 LAK-----DLTNL-----NSVN-----AGGTRI 1504
 Db 826 LNKQGAIFGVKQVIISGPGNIVFNEIGNVGIVHGAANSISFENASLGTSLFPGSTPL 885
 QY 1505 DEKGISFVDANGOA-KANTPVLISANGLDLGGKRSINIGAAVDDND---AYNFKQFNEVAK 1560
 Db 886 DVLTIKSTVCGTVDFNPNAPIVVVSIGID---SWINNGQIIGDKKNIILALSISDSDSITV 941
 QY 1561 TVNNL-----NQNSNGASLPFVVYTDANGKPIGDTGCKPQKAIKGADGKYYHAN-ANGVP 1614
 Db 942 NANTLYSGIRTTKNNQG-----TVTLSSGMPNN-----PGTIYGLGLENGSP 983
 QY 1615 VDKDGPITDADLANLAAGKPLDAGHQVVASLGGNSDAITLT-----NIKSTLP 1665
 Db 984 KLQVTFITDDYNNLSGLIANNVTI-----NDIVTLTTGGIAGTDFDAKITLG 1030
 QY 1666 QIDTPNTGNANAGQAQSL-----PSLSAAQQSNAASYKDLVNLGVFNLTQNHNOYDFVKAYD 1721
 Db 1031 SVN---GNAVRFVDFSTFSDPRSMIVATQANKGT---VTYLGNALVSNIGSLD-----T 1078
 QY 1722 TVNFVNGTGDADITSVSADGTMNSI-----TVNTALANTDDDDGNVLKAKGKFKYA 1773


```
Db 1079 PVASVRFETND-----SGAGLQGIYNSIDFGTYNLTIL-----NSNVILGGGTAINGE 1129
Qy 1774 DDLMPGSLKA-GKSASDAKPTGLSLVNPAGKSTGDAVALNNLSKAVFKSDGTGTTT 1832
Db 1130 IDLLTNLIFANGISTWGDNTSISTTL---NVSSNGIQGVIAEDAQ---VNAITTGTTI 1184
Qy 1833 TVSSDGIS-----IQGKNSSITLSKDLNVLGKGVISNVGKPKDQDAANV-QOL 1881
Db 1185 KIQDNANAFSGTQAYTLIQGARFNGTLAGPNAFVGTGNSIFVYELIRDSNQDVLTRT 1244
Qy 1882 NEVRNLL--GLGNAGNADGNQVNIADIKKDPNCGSSNSTVIAKAGTVLGGKNNDEK 1939
Db 1245 NDVLNVTTAVGNSAIAAPGVQNSIRCSLESTNTAAYNNMLLAK-----DPSS 1293
Qy 1940 LATGCVQGVKDGNGANGDLSNVVVKTKQDKGSKKALLATYNAAGQTNVLTNNPAEADRI 1999
Db 1294 VATFVGAIAITDTSAAVTTVNLNDTKQTD-----LLS--NRLGTLRLSLN--AETSD-- 1341
Qy 2000 NEQGRFRFHVNDGNEQPVVQG---RNGIDSSASGKHSVAIGFQAKADGEAAVAIGROTQA 2056
Db 1342 -VAGSATGAVSGDEAEVSYGVWAKPEFYNAEQDKKGGIAGYKAKTTG---VVVGLDPLA 1397
Qy 2057 GNOQ---STAIGNAQATGDOSIAITGTVNVTKHSGAIGDPSTVKADNSYSVGNNNQFID 2113
Db 1398 SDNLMIAGAAIGITTKDIKHQYKKGDKTDINGLSFSLYGSQOLVK--NFFAQN----- 1449
Qy 2114 ATQTDVFGVGNNTVETESVALGNSAISAGTHAGTQAKSDG---TAGTGTTAGATGT 2170
Db 1450 ---AIFTL--NKVKSQRFFESGKWSKQIAAGNYDNMTFGNLIFGYDYNAMPNVL 1503
Qy 2171 VKGFAGOTAVGAVS-----VGASGAERIQNVAAGEVSATSTDAVNGSOLYKATOGIANA 2225
Db 1504 VTPMAGLSYLSKNENYKGTGTIVANKRINSKFSDRV-----DLIVGAKVAGSTVNIITDI 1558
Qy 2256 TNEIDHRTQ--NEKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQGAVALGSLK 2284
Db 1559 V--IYPEHSFVHVHKNLSQSNL-----DQGTAPFISQP 1594
Qy 2285 SDNGQWFKINGSADTQGHVGAAGVAGPHF 2314
Db 1595 DRTAKTSYNIGLSANIKSDAKMEYGIYDF 1624

RESULT 7
OMP_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RL evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
```

```
RT *Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16353; CAA34403.1; -.
KW Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1333 32 KDA BETA PEPTIDE.
FT CHAIN 1334 1654 POLY-THR.
FT DOMAIN 1181 1188
SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 3.8%; Score 444.5; DB 1; Length 1654;
Best Local Similarity 21.8%; Pred. No. 5.6e-09;
Matches 409; Conservative 182; Mismatches 728; Indels 561; Gaps 88;

Qy 662 ITPAGENGISINDIAKGVKVGIDPINGL---TTPKLTVGSDK---GKTQLVIEQVAS 715
Db 72 ITANANINGINFNPAGS-----FNGLLNTANNLAVTVSEDTTLGFTINNVHNAHS 122
Qy 716 GNDTKNIIRGLSPITLPSITNAGGVRTTBOGNTITSDKSKAASIGDILNTGFNLKNSN 775
Db 123 ENLTLNACKTLITIGQVYNAAQAATKNAQNVVQ-----FNGAAIDNND- 168
Qy 776 SVGFVSYNTVDFIDGNATTAKVYDETQNTSKVYDVNVDEKTIELTGD-----NGKTN 830
Db 169 -----LKGVGRIIDFGAPATLVFNLANPTT-----QKAPLILGDNAVANGVNG 212
Qy 831 KIGVKTITLTITN---ANGKATNFSTTDNDALVNAKDAENLNTLAKETHT---KCTA 883
Db 213 TLNTNGFIQVSNKSFATVKAINTA--DQGLIIFND-ANNANTLNLOAGGTTINFTGTD 269
Qy 884 DTALQTEKVKKDGATDDETTITVGKDTQNGK--TVNTLKLKGENGLTVATNKDGTVTFCIN 942
Db 270 GTG-RLVLLSKHAAATNFNITGSLGNLKGVIETNFVAVDQ--LTANAGAANAV-IGTN 325
Qy 943 TQSLKAGDSTTLNKDGLSINKPASNEOIQGADGVKFAKVD-----KGNSTGIDGTSR 997
Db 326 NCAGRAAG-----FVVSVDNGKVATIDGGVYAKDMVYQSANATGQ 365
Qy 998 ITKQIGFTGANGSLDTPKPLHTKDKLVGVEITNTGINAGGKKTINIQSGDITONSND 1057
Db 366 VNFRIHVDVGADGT-----TAFKTAASKVT-----ITQDSNF 397
Qy 1058 AVT--GGRVYDLKTELESKINSAAKTAQNSLHEFSVADEQGNHFT--VSNPYSSYDTSKT 1113
Db 398 GNTDFGNLAAQIKVP-----NAITLT-----GN-FTGDASN-----GNT 431
Qy 1114 SDVITF-----AGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVGN-- 1155
Db 432 AGVITFDANGTLESASADANAVTNNTAIDASG---AGVVQLSGTHAAELRLGNAGSIF 488
Qy 1156 --KGIVIDSK-----DQONTITGLSNTLANTVNDGAGHALSQ-GLA 1194
Db 489 KLADGTVINGKVNQALVGGALAAAGTITLDGSATITG-----DIGNAGGAALQRIILA 542
Qy 1195 NDTDKTRAASIGDVLNAG---FNLOGNCEAVDFVSTYD--TVDFIDGNATTAKVYDDTS 1249
Db 543 NDAKKTTLTGGANIITGAGGGTIDQANGGTIKLTSTQNNIVVDFDLATATDQTGVVDASS 602
```


QY 1250 KTSKVVYDVN-----VDNKIEVTSKGLGKVTTLTKTSANGNA--TKESAADGDAL 1300
DB 603 LFNQATLTIGKIGTANNNKTL--GQFNIGSSKTVL-----SNGVAINELVINGNDGAV 655
QY 1301 VRASDIATHLNTLAGDIOTAKGASQASSASVVDADGNKVIYDSTDKKYQVNDKGVDPK 1360
DB 656 QFHD--TYLIT-----RITNAAGQ-----GKLIENPV-----VNN 684
QY 1361 NKEVAKDLVAQAQTPDGTGLAQMNVKSVINKQVNDANKKQGINEDNAFIKGLENAAKDT 1420
DB 685 GTTLAAGTNLGSATNP--LAEINFGS-----KGWVND----- 714
QY 1421 KTKNAAVTVGD--LNAVAQTPLETFAGDTGTTAKKLGETLTIK---GQOTDTNKLTDNNIGV 1476
DB 715 ----TVLNVGEGVNYATNITTTDANVGSVFVFNAGGTNIVSGTVGG-----QGNKFT 764
QY 1477 VAGTGGFTVKLAKDLTNLSNVNAGGTRIDEKGISFVDANGQAKANTPVLSANGLDLGGKR 1536
DB 765 VALENGTIVKFL-----GNATFNGNTTIAANSTLQIGGNY 799
QY 1537 ISNIGAAVDDNDVAFNKQFNEVAKTVNN-----LNNQNSG----- 1572
DB 800 TADCVASADGTGIVEFVNTGPIVTLNKQAAPVNALKQITVSGPGNVVINEIGNAGNHG 859
QY 1573 -----ASLPFVVDANGKPIGTGDKPKQKAIKGADGKYHANANG--VP----- 1614
DB 860 AVTDITAFENSILGAVFLPRGIPENDAGNTMPLTIKSTVG---NKTAKGFDVPSVVVLG 916
QY 1615 ---VVDKGPITDADKLANLA-----AHGKPLDAGHQVVASLGGNSDAITLTNIKT 1663
DB 917 VDSVITADGQVIGQNNVGLGSGNGIILVNATTIYAG---ISTLNNQGTIVLSG----- 969
QY 1664 LPQIDTPNTGNAGQAQSLPSLSAAQSVKVDLVNFGFNLOTNHNQVDFVK----- 1718
DB 970 ---GVNPTGTVYGLG---TGIGASKFKQVTFDTTDYNNLGNLIATINATINDGVVITGG 1022
QY 1719 ---AYD---TVNFVNGTGADITSVRSADGTMSTNTVNTALATDDGNNVLKAKDGKPY 1771
DB 1023 IAGIGFDGKITLGSVNGNG---NVRFADGILSN---STSMIGTTKANNGTVYILGNAPV 1075
QY 1772 K--ADDLMPGSLKAGKASDAKT-----PTGLSLVNPVNAKGSTGDAVALNN 1817
DB 1076 GNIGSDTPVASVRTGSDSGAGLOGNIYSQVIDPCTYNLGVNSNIILG---GGTIAING 1133
QY 1818 LSKAVFKSKDGTITVTSSDGSIGQKDNSSI--TLSKDGLNVGKVI---SNVKGKTKD 1872
DB 1134 -----KIDLVNTLTTFASGTSTWG--NNTSIETTLTLANGNIGHIVILEGAQVNTTTG 1185
QY 1873 TDAANVQQLNEVRNLLGLGNAGDNADGQ-----NA--NANFSGTQVTLTIOGGARFNGTLGSPNFVATGSGNRFVN 1233
DB 1186 TTTIKVQD-----NA--NANFSGTQVTLTIOGGARFNGTLGSPNFVATGSGNRFVN 1233
QY 1920 RTVIRAGTVLGGKGNNDTEKLTATGGVQGVQVDRD-----GNANGDLSNV--VWKTQKDG 1971
DB 1234 YSLIRA-----ANQDYVITFTNAENVVTVNDIANSFPGAPGVQDVQNVTFVNATNTRAA 1286
QY 1972 KKALLATYNAAGQTNY---LITNPAAIDRINEQIRFPHVNDGQ----- 2014
DB 1287 YNNLLAKNSANSANFVGAIVTDSAAITNVQDLDAKDIOAQLGNRGLALRYLGTPTAE 1346
QY 2015 -----EPVVOGRNIDSSAG-----KHSVATGFOKADGEAAVAIG 2051
DB 1347 MAGPEAGAIISAAVAAGDEAIDNVAYGIWAKPFTYDAHQSKKGLAGYKAKTTG---VWIG 1403
QY 2052 ROTOAG-----NOSIAGDQAATGQDSIAIGTGNVVTG--KHSGAIG 2092
DB 1404 LOTLANDNLMIGAAITGTTDKHODYKGDXTDVNGSFSLYGAQQLVKNFPAQGSALF 1463
QY 2093 DPFSTKADNSY-----SVGNNNQFIDATQDVFVGNNITVTSNVALGNSAISAGTHA 2148
DB 1464 SLNQVKNKSQRYFFDANGNWSQIAAGHYDNMTFGNLTGVGDYNAMQGVLPVPMAGL-- 1521
QY 2149 GTQAKKSDGTAGTITTTAGATGTVKVG--FAGQTAVGAVSVGASGAERRIONVAAGEVSATST 2207

DB 1522 -SYLKSSDSENYKETGTTTAVANKOVNSKESDRTDL-IVCAKAVAGSTMNITDLAV-----Y 1572
QY 2208 DAVNGSQLKATQGIATNATNELDHRIT-----HQENKANAGISSAMAMASMPQAYIP 2259
DB 1573 PEVHAFVHVHKVTCRLSKTOSVLDGQVTPCINQPDRTTKTSYNLGLSASIRSADKMEYGIG 1632
QY 2260 GRSMVTGGIATHNGOGAVAV 2279
DB 1633 YDAQISSKYTAHQGTCLKVRV 1652
RESULT 8
OMPB_RICTY
ID OMPB_RICTY STANDARD; PRT: 1645 AA.
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON;
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RT surface layer protein of Rickettsia typhi.";
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=WILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: Identification of an avirulent
RT mutant deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L04661; AAB48987.1;
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).

FT	CONFLICT	657	657	H -> N (IN REF. 2).
FT	CONFLICT	842	842	V -> I (IN REF. 2).
FT	CONFLICT	1071	1071	G -> A (IN REF. 2).
FT	CONFLICT	1306	1306	G -> S (IN REF. 2).
SQ	SEQUENCE	1645	AA; 169698 MW; OCB5641C7EB165EE CRC64;	

Query Match			
Best Local Similarity 3.7%; Score 438.5; DB 1; Length 1645;			
Matches 390; Conservative 209; Mismatches 570; Indels 599; Gaps			
QY	725	GLSPITLP	SITNACGVRVTTGGNTIISDEDKSAAASGIDILNTGPNLKN-NSNSVGFFVSTY 783
DB	59	GAGYNLP	VPATNS--VITANSNAIT-----FNTGPNLSL-FLDTA 97
QY	784	NTVDFIDGN	ATTAKVYVDETNQTSKVYDVNDEKTIELTGDNCKTNKIGVKTTLTTTN 843
DB	98	NTLAVTIN	EMTITLGFVNTVTKGPFNFITGAG-KSLTIGH-----GITAQAAATK 149
QY	844	ANGKATN	-FSTTDNDALVNAKDAETAEMLNTLAKETHTKTADTALQTKVKKDGATDDET 902
DB	150	S---	AQNVSVKNVAGAAINDNLS-----GVGSIDFTAAPSVLLENL-INFTQEAP 197
QY	903	ITVGKDG	TQNGKVTNVLKLGKGNLTIVATNKDGTVTFGINTQSGULKAGDSTTLN---KDG 959
DB	198	LTLG---	DNAKTVN---GANGILINIIN--GPKVSKDTFAGIK-----TINIGNQG 241
QY	960	L---	SIKNPASNEQIOGADGVKFAKVDKGNSTGIDGTSRI-----TKDQIGFTCA 1009
DB	242	LMFTPTD	AAALNLOGGNTINF-----NGRDGTGLVLVSKNGNATEFNVTGSL 292
QY	1010	GLSDTTK	PHLTOKDLKVGVEITNTGINAGGKKITN-----TQSG 1049
DB	293	GG-----	NLK-GVIEFDTTA--AAGKLIANGGAANAVIGTDNGAGRAAGFIVSV 338
QY	1050	DIQNSND	AVTGRVYDLKTELESKINSAAKTAQNSLH-----EFSVADEQ----- 1095
DB	339	D---	NGNAATISQVYAKDIVIQSANAGQVTFEHLVDVGLGGKTNEFTADSKVLIITENA 395
QY	1096	-----GN	---HFTYSNPF-----SSYDTSKTSDVITFACENGITTKVNGK----- 1132
DB	396	SPGSTDF	GNLVAQIVVPPNNKILTGPIGDAKKNNGTAGVITF-NANGPLVSGNTDPNIV 454
QY	1133	-----VVRVG	IDQTKGLTTPKLVGNNN-----GKGIVIDSKDQQTINGLSNTLAN- 1179
DB	455	TNIKAIE	VEGAGIVQLSGIIGAELRGNAGSIFKLADGTVINGPNQNPVYN-NNALAAG 513
QY	1180	-----VTND	---GAGHALSQG--LANDPDKTPRAASIGDVL--NAG--FNIQGNCEA 1221
DB	514	SIOLDG	SAIITGDIGNCVANAALQDITLANDASKILTLSGANITICANAGGAIHFQANGGT 573
QY	1222	VDFVSYDT	--VDFIDGNATTAKVYDFTSKTSKVYDVNVDNKITIEV--SDKLLGVKTT 1278
DB	574	IQTSTQN	ILLVDF-DLDVTTDQTVVDASSLT-----NNQTLTINGSIGTIGANTK 624
QY	1279	TLTKTS	SANGNATKFSAADG--DALVKASDIAPHL--NTLAGDIQDTAKGASQASSASYVD 1334
DB	625	TLGRNV	SGSKTILNAGDAVAINELVWENDGSVHLTHNYL-ITKTINAQ----- 674
QY	1335	ADGNKVI	YDSTDKYYQVNDKGOVDKNKEVAKDKLVQAQPDQGTLAQMNVKYSVINK-EQ 1393
DB	675	---GKII	-----VAADPINTDTALADGT---NLGSAESPLSN 705
QY	1394	VNDANKQ	QGINEDNAFI--KGLENAAKDTTKNAAV-----TVGDNLAVA 1436
DB	706	IHFATK	--AANGDSILHICQGNVLVANNITTTDANVWSLHFRSGSTSVSGVGGQGLK 763
QY	1437	QTPLT	FAGDTGTTAKKLGETLTIKGG-----QDTNKLTD-----NNIGVV--AGT 1480
DB	764	LNNLIL	-DNGTIVKFLGD-ITFNGKTECKSILQISSNYITDHIESADNTGTLEFVNT 820
QY	1481	DGFTVK	LAKD-----LTMLSNVNAGGTRIDPKGISFVDANGQAKANTPVLSANGLDLGG 1534

RN SEQUENCE FROM N.A.
 RC STRAIN=NCDO 151;
 RX MEDLINE=92381481; PubMed=1512565;
 RA Holck A., Naes H.;
 RT "Cloning, sequencing and expression of the gene encoding the cell-
 RT envelope-associated proteinase from *Lactobacillus paracasei* subsp.
 RT *paracasei* NCDO 151.";
 RL J. Gen. Microbiol. 138:1353-1364(1992).
 RN [2]
 RP SEQUENCE OF 189-196.
 RX MEDLINE=92226694; PubMed=1564442;
 RA Naes H., Nissen-Meyer J.;
 RT "Purification and N-terminal amino acid sequence determination of the
 RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.
 RT *paracasei*.";
 RL J. Gen. Microbiol. 138:313-318(1992).
 CC -!- FUNCTION: PROTEINASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
 CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
 CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
 CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS.
 CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
 CC INSULIN B-CHAIN.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M83946; AAA25248.1; -;
 CC PIR; B44858; B44858.
 CC DR HSSP; Q99405; 1MPT.
 CC DR MEROPS; S08.019; -;
 CC InterPro; IPR000209; -;
 CC InterPro; IPR001899; -;
 CC Pfam; PF00746; Gram_pos_anchor; 1.
 CC Pfam; PF00082; Peptidase_S8; 3.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASE_ASP; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 CC KW Hydrolase; Serine protease; Cell wall; Zymogen; signal;
 CC Transmembrane.
 CC FT SIGNAL 1 33 POTENTIAL.
 CC FT PROPEP 34 187 POTENTIAL.
 CC FT CHAIN 188 1902 PII-TYPE PROTEINASE.
 CC FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).
 CC FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).
 CC FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 CC FT PROTEINS.
 CC SEQUENCE 1902 AA; 200253 MW; D8C9F38CEE5DA582 CRC64;
 Query Match 3.5%; Score 407.5; DB 1; Length 1902;
 Best Local Similarity 21.1%; Pred. No. 1.4e-07;
 Matches 450; Conservative 262; Mismatches 856; Indels 567; Gaps 109;
 QY 400 GLVKVLAQN--LSGLETVSTKNTLA---SEKVTVGSGNNTAELOSGGLTFTPTTNASTD 453
 DB 7 GLSILAGTVAGALAVLPVGEIQAKAASISQPKVSSLANTVRAAKAQAATDTTAATTN 66

QY 454 KTVYGTDLKFTDNSNTALEDDTTRITKDKIGFSNKAGTVDENKPYLDKDKLVGNSTLNN 513
 DB 67 QAI-----ATQLAAKGIDY-NKLNKVVQQDTYVDVIVQMSAAPASEN 107
 QY 514 GGLTVNNTGGSNKOIQVGADGKIFADVNVNVAAKAFGTTRITEEIGFADADGKY--- 570
 DB 108 GTL---RTDYSSTAEIQOETNKVIAQAQSV-----KAAVEQVTOQTAG--ESYGYVNG 156
 QY 571 -DKKSPYLDKKOL-QVGGVKIT-----KDSGINA-----GQOKISNVK 606
 DB 157 FSTKVRVVDIPKIKQIAGVKVTVTLAKVYPTDAKANSMANVQAVSNKYKGGEGTVSVVI 216
 QY 607 DAFDDTDAVYTKOLKOVQDADGALQSFSIRDEKGOEFTISNLYSNGNTPTTETITFAG 666
 DB 217 DTGIDP---THKDMR-LSDDKDVKLTKYDV--EKFTDTAKHGRYFTSKVPYGEN---YA 266
 QY 667 ENGISISNDIAKVKVGVGIDPINGLITPKLTGVSDDKDKTQLVIEQVAVSNDNDKNIIRGL 726
 DB 267 DNNDTITDDT-----VDEQHMHVAGI-IGAN-----GTGDDPTKSYVG 305
 QY 727 SP-----TLPSITNAGGVRTTEQNTITTSDEKSKAASIGDILNTGPNLNKNSNS----- 776
 DB 306 APAQALLAMKVFTNSDTSATGTSATLVSALEDSAKIGA--DVLNMSLGSDSGNOTLEDPE 363
 QY 777 VGFVSTYNTVDFIDGNATTAKVTVYDETNTQTSKYTVYDYNVDEKTIELTDNGKTNKIGVKT 836
 DB 364 IAAVQANNE-----SGTAAVISAGNSGTSGSKGVNKDYGLQ---DNEMVGTGPG--T 412
 QY 837 TTUTTTNANGKATNFSTDDNDALVNAKDAENLTLAKEIHTTKGTADTALQTFKVKKD- 895
 DB 413 SRGATTVASAENTDV-ISAQVTTIDGKDLQLPETITQLSNDEFTGSPDQ--KKFYVVKDA 469
 QY 896 -----GATDEE-----ITVCKDGTQN-GKVTNTLLKGENGLTVATNKDGTV----- 937
 DB 470 SGDSLKGAAADYADAKGKIATVIRGELNPAQKQVAAQAAGL-IIVNDGTATPLTS 528
 QY 938 -----TFGINTQSLKAGDSTTLN-KDGLSIK-----NPA 966
 DB 529 IRLTTFTTFFGLSKKTQKLVDMVTAHPDDSLGVKIALTLPLNQKYTEDKMSDFTSYGPV 588
 QY 967 SNEQIQ--VCADGVKPAKVDKGNSSGIDGTSRTIKDQIGFTGANGSL----- 1012
 DB 589 SNLSFKPDITAPGGINWSTQNNNGYTNMSTGSMASP-----FIAGSQALLKQALNNKNPF 644
 QY 1013 -----DTTKPHLTDKLKVGEVEITNTGINAGKKITNIQSGDITQNSNDVAT-----GGR 1063
 DB 645 YADYKQLKGTALTDFELKTVEMN-TAQPIN-----DI--NYNVIVSPRROGAG 689
 QY 1064 VYDLKTELES-KINSAAKTAQN-----SLHEFSVADBEQ-----GNHFTVSNPYSSYDTSK 1112
 DB 690 LVDVKAADALEKPNSTVVAENGYPAVELKDTSTDKTFLFTNRTTHELTQYQMSNTD 749
 QY 1113 TSDVITFAGE--NGI---TTKVNKGVVRVIGIDQT----- 1141
 DB 750 TNAVYTSATDPNSGVLYDKKIDGAAIKAGSDITVPAGKTAQIEFTLSLPKSFQDQOQVEG 809
 QY 1142 -----KGLTTPKLTAV-----GN-NNGKGIYDSKDG-QNTITGLSNTLANVTNDGAGHA 1188
 DB 810 FLNFKSDGSRLLNLPYMGFFGDWNDGK--IVDSLUNGITYSPAGGNYCTVPLLTNKNTHQ 867
 QY 1189 LSQGLANDTDKTRAASIGDVLNAGFNLCQNGEAVDFVSTYDVTDFIDGNATTAKVTVYDDT 1248
 DB 868 YGGMVTADGKQTV-----DDQATAFSS-----DKNA-----LYNDI 900
 QY 1249 SKTSKVYVDVNVNKNKTIETVTSDDKLGKVTTL-----TKTSANGNATKF-----SAADG 1297
 DB 901 SMOYLYLR--NISNVQVDIILDGQ--GNKVTTLSSSTNQTKTYDAHSOKYIYYNAPAWDG 956
 QY 1298 DALVKASDIATHLNTLAGDIQTAKGASOASSASYVDADGNKVIYD---STDKKYQV-- 1352
 DB 957 -----TYDQROGNKRTADGSYTYRISGVPEGDKQVDFVPFKLDSKAPTVRH 1006


```
QY 1353 -----NDKGOVDKNKEVAKDLVAQAQTPDQTLAQNVKSVINKBOVDNANKKQGIN 1404
Db 1007 VALSAKTENGKQYYLTAE-AKDDL-----SGLDATKSVKTAINEVTNLDATFDAGT 1058
QY 1405 EDNAETKGLENAADTKTKNAATVVGDLNVAQOTPLTITAGDTG-----TAKKLGET---L 1457
Db 1059 TADGYTK-IETPLSDEQ-----AAALGNGDSEALYLTDNASNATNQDASVQKPGSTFSL 1113
QY 1458 TIKGGQTDNKLTDNNIGVAGTDG-----FTYKLAADLTNLSVNAVAGSTRIDEKGIS 1510
Db 1114 IVNGGGI-PDKISSTTGYEANTQGGTYTFSGTYPAAVD-----GYTDAQCKK 1162
QY 1511 FVDANGQAKANTPVLSANGLDLGGKRISNIGAAVDDNDVAFNKQFNE-----VAKTVNNL- 1565
Db 1163 H-DLNTTDAATNSTASMAVNTADYAAQVDLYADKAHTQLLKHFDTKVRLTAPTFTDLK 1221
QY 1566 -NNQNSGASLPFVVT---DANGKPIGTGDKPKAKAIGADGKYYHANANGVPV----- 1615
Db 1222 FNGSDQTSKATIKVTGTVSSDTKTVNVD-----TVAALDAQ---HHFSVDVPVNYGDN 1274
QY 1616 -----DKDG-----KPIIT---DADKLANLAHKGKPLDAGHVVASLGNSDAITLNI 1660
Db 1275 IKVTATDEGNTTTEBQKTISSYDPLKNAV-----FDQG-----VKFGAN-----EF 1319
QY 1661 KSTLPQIDTPNTGNANAGAOQSLSAQAQSNAAVSKVDLVNCGFNQTNH-NQVDF-VK 1718
Db 1320 NATSAKFDPKGTGATITGKVKHPTTTLQVQDKQISIKNDLTFSTLDTLGLQKPPFGV 1379
QY 1719 AYDVTNVFNGTGAD-ITSVRSADGTMSTNVTNLAATDDGDNVLI---KAKDGKPKKAD 1775
Db 1380 VGDITQ--NKTFQEAFTILDVAFTLSLSDSTDAVYTNDFQITGTATDNAQYLS-- 1435
QY 1776 LMPNGLKAGKSADAKTPTGLSLVNPAGK---STGDAVAL-----NNLSKAVFKSKDG 1828
Db 1436 LAINGSHVASQAD-----ININSGKPGHMAIDQPVKLLGKVLVAVATDSENN 1485
QY 1829 TTTTIVSSDGISIOCKDN-----SSITLSKDLGVLNCGKV--ISNVGKGTGD 1872
Db 1486 TTTTKIT---VYIEPKKTLAFTVTPSTTEPAKVTTLTANAAATGETVOYADGGKTYQD 1542
QY 1873 TDAANVQ-QLNEVRLLGLGNAGDNADGNQVNIADIKDP-----NSGSSSNRTVIKAGT 1927
Db 1543 VPAAGVTVTANGTFKFKSTDLYGNSPAVDYV-VTNIKADDPAQLOTAKQALTNLIASAK 1601
QY 1928 VLGGKGNND-----TEKLTAG--GVQGVDDKGNANGDLSNVVWVKTKDGSKKAL 1975
Db 1602 TLSASGKYDDATTTALAAATQKQATLQDQTDASVDSLGTGANRDLQT-----AINQL 1652
QY 1976 LATYNAAGOTNYLTNNPAEADIRNEQIRPFHVNDGQEPVQGRNGIDSSASKHVA 2035
Db 1653 AAKLPADKKTSL-----NQLOSVKALGTDGLNQTDPSGTG 1688
QY 2036 IGFOKADGEAAVATGROTAQGNOSIAIGDNAQTDGQSIAGTGNVTVGRHSGAIGDPS 2095
Db 1689 KTFEAAALDLVA-----QAQAGTQD---ADQLQA-----SLAKVLDLAVLAKLAEG----- 1730
QY 2096 TVKADNSVSGNNNOFIDATQTDVFGVGNNTVVTESNVALGSN-----SAIS 2143
Db 1731 -IKAATPAEVENAK---DAATGKTMYADIADTLTSGQASADASDKLAHLQALQSLKTKVA 1786
QY 2144 AGTHAGTCAKSGDGTAGTITTTAGATGTVKGFAGOTAVCAVSVG---ASGAERRIONVAAG 2200
Db 1787 AAVEAAKTAGKDDTTGTSGDKGGGTAPAPAGDTGDKDKDEGQSPSSGG-----NIPTK 1841
QY 2201 EVSATST---DAVNGSOLYKATQGIANNATNLDHR 2232
Db 1842 PATTTSTSTDDTTDRNGQHTSGKALPKTAETTER 1876

RESULT 10
CBPA_CLOCL
ID CBPA_CLOCL
AC P38058;
```

```
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULOSE BINDING PROTEIN A PRECURSOR.
GN CBPA.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN NCBI_TaxID=1493;
RP SEQUENCE FROM N.A.
RX MEDLINE-92228810; PubMed-1565642;
RA Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
RT binding protein A.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC -I- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE
CC ENZYMES.
CC -I- PTM: THE N-TERMINUS IS BLOCKED.
CC -I- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: M73817; AAA23218.1; -
PIR: A44140; A44140.
HSP: O06851; INBC.
InterPro: IPR001956; -
InterPro: IPR002102; -
Pfam: PF00942; CBD.3; 1.
Pfam: PF00963; Cohesin; 9.
Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
KW SIGNAL
FT CHAIN 1 28
FT DOMAIN 29 189
FT SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F71AF1A CRC64;
SQ
Query Match 3.5%; Score 406; DB 1; Length 1848;
Best Local Similarity 20.6%; Pred. No. 1.5e-07;
Matches 452; Conservative 277; Mismatches 800; Indels 668; Gaps 111;
QY 31 SSSTAGQVGS-SPVRLTRVATLALV-----IGATLNGSAY 66
Db 39 NSNKSQNTSIPPIIKITNTSDSLNLDVKNVRYVYSDGTQGTGFWCDHAGALL-GNSY 97
QY 67 AONSKIAFGTTGNNDNASNEASIAIG---SLAKAHANQAIAGG---SKPDPNRN----- 117
Db 98 VDNSTKVNTANFVKETASPTSTYDYVEGFGAGRATLKKGFITIQGRITKSDSNYQT 157
QY 118 -----QAANQKAGSHAKKESI--AIGDVLAEQASTAIGSDDLVLDNRNSTNS 164
Db 158 NDYSFDASSSTPVVNVKVTYIGGKVLGTAPGPDV-----PSSIINPTSATFKNVTKQ 212
QY 165 KYPNGLLSTLIQNTLVLRQIRDSNGSKYRRTAAEGHASTAVGA--MAYAKGHEANFGT 222
Db 213 A---DVKTMTFLNGNTFKTITDANG-----TALNASTDYSVSGNDVTLISKAYLAKOSVG 263
QY 223 RSTAEKNYSL-----AVGLT---AKAEKGYTIAIGSNAQAINTYCALAL- 262
Db 264 TTTLNFESAGNPQKLVITVVDTTPVEAVTATIGKVQVNAGETVAVPNVLTVPKPAAGLATI 323
QY 263 -----GADTRVDLDYGIALGYCSQILNNNNNNNNNNKAYVEPEGNSNIK-----SS 306
Db 324 ELPTFDSASLEV-----VSTAGDIVLNPVNSSTV-----SGSTIKLLFLDLDLGSQ 373
QY 307 KATGNGLFSGSSTIKRKIIN-----VGAGYEDTDVAVNVAQLK-----AVENLAKRQITF 356
```


Db 374 LITKDGVEA--TITPKAKAITGTTAKVTSVKLAGTVPVGDALQOEKPCAVN---PGTVTI 428
QY 357 KGDNDGTGVKKILGETLTITIKGSETQADKLTDDNNNIGVWTDNNITGLKVLAKNLISGLETVS 416
Db 429 NPIDNRMQI--SVG-TATVKAIEAAPPV-----LTSVPSTG-----IA 465
QY 417 TKNLTASEKVTGSGNNTAELOSGLTFTPTTINASTDKVTYGTGDKLFTDNSNTA-----L 472
Db 466 TABAQVSFDATL---LEVASVTAGDIVLAPTVNFS-----YTVNGVVIKLLFL 510
QY 473 EDT---TRITKD---KIGFSNKAAGTVDENKPYLDKDKLVGNSTLNNGLTIVNNTIGGS 525
Db 511 DOTLGSOLSKDGVFTINFAKAVTSTVTP-----VTVSGT-----548
QY 526 NKQIQVAGADIKFADYVNVVNSNAKFGTTRITEE--EIGFADADGVKKSPYLDKQLQ 593
Db 549 ----PVFADGTL---AEVQSKTAAGSVTINIGDPILEPTISPTATFDKAPADVATTMT 601
QY 584 VGGVKITKDSGINAGDQKIS-NVKDATDQTDATVYKQLKQVQDADGALQSFIRBEKQ 642
Db 602 LNYTTFNGITGLTSDYSISGNVVKIS-----QAYLAKQPVG-- 638
QY 643 EFTISLNSNGNTPNFTITFAGENGISINDIAKGVKVG-----IDPINGLTTP 694
Db 639 DLTLTFNSGNKTATAKLVSIKDPKIVTATVGTATVAGETVAVPVTLNVSGISTA 698
QY 695 KLVGSDKDGKQTLVIEQVAGSNDTKNIIRGLSPITLPSITNAGGVRTTEQGNITSDDEK 754
Db 699 ELQLSFD---ATLEVVISITAG---DIVLNPVSFSSVNV-----GSTI-----736
QY 755 SKAASIGDILNCGFNKLNKNSNVGFSTVNTVDYFDIGNATTAKVTYDENQTSKVTYDVN 814
Db 737 -KLLFDDDTLGSOLSKD-----GVFATINF-RAKSVTSTVTPVKVSGPPVFADGT 786
QY 815 VDEKTIETLGDNGKTKIG-VKTTILITITNA---NGKATNFSTTDNALVNAKDIAENLN 870
Db 787 LAELSVETVAGSVTINAIGPVKIVTATVGTATVKSETVAVPV-----LSNVPGIA----838
QY 871 TLAKETHTTKGADTALQFVKKKGDATDETTIVGK-----DGTQNGKTVNWLKL 921
Db 839 -----TAELQLSFDATLLEVA---SITVGDIVLNPVSFSSVNVSGTIKLLFL 883
QY 922 KGENGLTVATNKDG---TVTFGINTOSGLKAGDSTTLNKGSLIKNPASNEQIQVADGV 978
Db 884 DDTPLG-SOLISKDGLVATINFKAKTVTS-----TVPVAVSGPPVFADGT 938
QY 979 KFAKVDKSGNSTGIDGTSRITKDIQIGFTGANGSLDTTKPHLTDKLKVGE-----VEITN 1033
Db 929 -LAEL---QSKTVAGSVTIEPSQ-----PVKIVTATVGTATVKSETVAVPVTLN 975
QY 1034 -----TGINAGKKITNTQSGDI-----TQNSNDVATGGRVY-----DLKTELES 1073
Db 976 VPGIATAELQVGFADTALLEVASITVGDIVLNPVSFSSVNVSGTIKLLFDDTLGSOLIS 1035
QY 1074 K-----INSAKTAQNSLHEFSVADEQGNHFTVSNPYSSYDTSKTSVDITFAG-----EN 1123
Db 1036 KQGLVATINFKAKTV-----TSKVTTPVAVSGTPVFAD 1068
QY 1124 GITKVNKGWV--RVGIDOTKGLTTPKLTGVNNNGKGVIVDSKDGQNTITGLSNTLANVT 1181
Db 1069 GTLAEINMKTVAGSVTIEPSQPKVTTATVG-----TATVKSETVAVPV--TILSNVP 1119
QY 1182 NDGAGHALSQGLANDTKTRAASIGD-VLNAGFNLOG--NGEAVDFVSTYDVT--DFIDG 1236
Db 1120 GIATAE-LQVGFADTALLEVASITVGDIVLNPVSFSSVNVSGTIKLLFDDTLGSOLIS 1178
QY 1237 NATTAKVTYDDTSTKTVYDVNVNOKTI-----EVTSDKGLGVKVTTL 1280
Db 1179 DQVLATINFKAKTVTSKVTTPVAVSGTPVFADGTLAEKLYETVAGSVTIEPSQPKVTTA 1238
QY 1281 TKTSANGNATKFSAAD-----GDALVKASDIATHL---NTLAGDITOTAKGASOAS 1328
Db 1239 TVGTATGKVGETVAVPVTLISNVPGIATAEVQVGFADTALLEVASITAGDIV----LNPSVN 1294

QY 1329 SASYVDADGNKVIYDSTDKKYQVNDKGQVKNKEVAKDKLVAQAQTPDGTLAQMNKSV 1388
Db 1295 FSSWNGSTIKILF-----LDDTLGSOLISKDGVFTATNFKI- 1331
QY 1389 INKEQVNDANKKOGINEDNAFIKG-LENAAKDTKTKNAAVTVGDLNAYAOFTPLTFAGDTG 1447
Db 1332 --KAVPSTGTPPAISGTPVFADGTLAEVOYKTVAGSVTIAAADIKAVKAT-----VG 1382
QY 1448 TTAKKILGETLITIKGGQTDTKNKLDDNNIGVAGTGDGTVKLAKDLTLN--SVNAGGTRID 1505
Db 1383 TATGKAGDIVAV-----PVTLSNVSGIATVE---LQLSFDATLLEVASITAGDIVLN 1431
QY 1506 EKGISFVDANGQAKANTPVLSANGLD--LGGKRISNTIGAADDNDVAFNFKOFNEVAKTVN 1563
Db 1432 P-----SVNFSSVWNGSTIKILFDDTLGSOLISKDGVFA---TVNFK---VKSTAT 1477
QY 1564 N-----LNNQNSNG--ASLP-FVVTDANGRPINTGDKPKQA-- 1597
Db 1478 NSAVTPTVTVSGTPVFADGTLAEELKSESAAGRLTILPTVIIIVDSTVAPTAVTFQKANOADA 1537
QY 1598 --IKGADGKYVHANANGVPVDKGPITTDADKLANAAHKGPLDAGHVVASLGGNSDAI 1655
Db 1538 AITWTLNGTFTSAIKNGTATLVKGTDTYVSENVVTISK-----AYLAKQGTV 1585
QY 1656 TL-----TNIKSTLPQIDTPTGNANAGQAQSLPSLSAAQQAQSNAAVKDVLNVGFN 1706
Db 1586 TLEFVFDKGSNAKVVAVAKIIVNS-----TITPVATFEKTAQAQADV---VT 1633
QY 1707 LQTNHNOVDFVKAYDIVNFVNGTGCADITSVRSADGTM--SNITVNTALATDDGCVLIK 1764
Db 1634 MSUNGNTFTSAIK-----NGT---TTLVKGTDYITISGTVTISKAYLATLADGSAATLE 1682
QY 1765 --AKDGKFKY-----ADLMPNGSLKAGK---SASDAKPTGLSLVNPVNAKGST 1809
Db 1683 FVFNQOASAKRLTIVPAVDPVVTDFAVKIDKVSAAAGSTVKVP--VSLINVS-----1734
QY 1810 GDAVALNLSKAVFK-SKQGTMTTIV--SSDGIISQKDNSSITLSKDLNNGVKVINSVG 1867
Db 1735 ---KGVNVCVAEYKISFSDSVLTIVGTAGTSIK---NPAVNFS-----1772
QY 1868 KGTKDDTDAANVOQLNEVRNLLGLGNAGN-----DNADGNQVNIAD-----IKKDPNSG 1915
Db 1773 -----SOLN-----GNTILFDFDNTIGNELITADGQFATIEFKVNAA 1810
QY 1916 SSSNRTV-IKAGTVLGGKGNNDTE--KLAT--GGVQV 1947
Db 1811 ATSGTTAEYKVAISSFADASLAEITKVAIVNGSVKV 1847

RESULT 11

HLVA_SERMA
ID HLVA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HEMOLYSIN PRECURSOR.
GN SHLA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC STRAIN=SN8;
RX MEDLINE=88257037; PubMed=3290200;
RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia marcescens";
RL J. Bacteriol. 170:3177-3188(1988).
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY

QY 2040 AKADGEANAIGR-----QTOAGNQSTA-TGDNAQTGDSIAIGTGNVYTKHSGA 2090
 Db 1437 PQDVTTCVAVSFAEGKVTLPATPAGEKPGQPLWDRGART-----VGGAVKDSITGP 1488
 QY 2091 IGDPSTVKADNSVSGNNQFIDATQDVFVGNNITVTESVAL--GNSNSAISAGTHA 2148
 Db 1489 AGRQGHUKV--NADVNNN-----AVGEQNSIAKNGVALQVGGQTQLTGGEIR 1535
 QY 2149 GTOAKSDG-----TACTTTTAGAT--GTVKGFAGQTAAGVAVSVCASGAER 2192
 Db 1536 SQQKVELGGSQVSDQVNGQRQYGGGRVDAATVGLLGAARKQSVAGNVP----- 1587
 QY 2193 RIONVAAGEVSATSDAVNG 2212
 Db 1588 ----FASGHASTQOADAAG 1603

RESULT 12

WAPA_BACSU STANDARD; PRT: 2334 AA.
 AC Q07833;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE WALL-ASSOCIATED PROTEIN PRECURSOR.
 GN WAPA OR N17G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of a
 RT Bacillus subtilis 168: evidence for processing of the product of a
 RT gene encoding a 258 kDa precursor two-domain ligand-binding
 RT protein."
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wapa loci."
 RL Microbiology 141:337-343(1995).
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC -!- MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES.
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 CC SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L05634; AAA22883.1;
 CC DR EMBL: D31856; BAA06656.1;
 CC DR EMBL: D29985; BAA06260.1;
 CC DR EMBL: D83026; BAA11683.1;
 CC DR EMBL: Z99124; CAB15959.1;

DR PIR: S32920; S32920.
 DR Subtilist; BG10797; wapa.
 DR Pfam; PF02018; CBP_6; 1.
 KW Cell wall; Repeat; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 2334
 FT DOMAIN 504 869
 FT REPEAT 504 605
 FT REPEAT 636 736
 FT REPEAT 769 869
 FT DOMAIN 1021 2139
 FT REPEAT 1021 1040
 FT REPEAT 1042 1061
 FT REPEAT 1063 1082
 FT REPEAT 1083 1102
 FT REPEAT 1109 1128
 FT REPEAT 1129 1148
 FT REPEAT 1150 1169
 FT REPEAT 1174 1193
 FT REPEAT 1199 1218
 FT REPEAT 1219 1238
 FT REPEAT 1646 1665
 FT REPEAT 1667 1686
 FT REPEAT 1690 1709
 FT REPEAT 1711 1730
 FT REPEAT 1732 1751
 FT REPEAT 1753 1772
 FT REPEAT 1795 1814
 FT REPEAT 1820 1839
 FT REPEAT 1840 1859
 FT REPEAT 1861 1880
 FT REPEAT 1887 1906
 FT REPEAT 1908 1927
 FT REPEAT 1929 1948
 FT REPEAT 1969 1982
 FT REPEAT 1983 2002
 FT REPEAT 2008 2027
 FT REPEAT 2028 2047
 FT REPEAT 2051 2070
 FT REPEAT 2071 2090
 FT REPEAT 2093 2112
 FT REPEAT 2120 2139
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;

Query Match 3.4%; Score 403; DB 1; Length 2334;
 Best Local Similarity 20.5%; Pred. No. 2.5e-07;
 Matches 481; Conservative 290; Mismatches 872; Indels 700; Gaps 122;
 QY 385 LTQN--NNIGVVDNNTGLKVKLAKNLG--LETVSKN-LTASEKVTGSGNNTAELQS 439
 Db 258 MTDKDELSEGEVSVDKSVKLEKEEGYLLHLTADENWLKDPERYVPYSIDSTSLV 317
 QY 440 GGLTFT----PTTNASTDKTVYGDGLKFTDNSNTALEDTTRTKDKIGFSNAGTVDEN 495
 Db 318 SSOTFVMSAYPTTN-----YSASSQKWDANLKAYVL-----KTGYDK--TTGTN 360
 QY 496 KPYLDKDKL--VGNSTLNLNGLT--VNNTIGGSNKQIQVGADGKFKADVNVNVSNAKFG 552
 Db 361 YAFMKFNLPIONMTVTKATLKYVAHSYGTK-----ATGLMLDTVNSYDNKVTW 414
 QY 553 TTRITEEIEGFADAGDKVDKKSPLYDKKQLQ-----VGGVKITDKSGINAGDKISNVKDA 608
 Db 415 NTRPASKNICKAD-----VHKGQWASVDVTAVKSWNSGGANYG-----FKLH 457
 QY 609 TDDTDAVYKQLQVQDADGALQSFSIRDEKGOEFTISNLYSNGNTPNTFTTTFAGEN 668
 Db 458 TNGNGKEYWKLISSANSANKPYIEVTYTPKGNTPTI--KAYHNGDSTGYFD--ISWKKVE 515
 QY 669 GISISNDIAGKVKVGDTPING-----LTTPKLTVGSKDKGKQLVIEQVAS----- 715
 Db 516 G-----AKG-YKWI--YNGKEYQAISAGNVTWSWTKGKKIWP TSAEIAASKRYKUHL 565

QY 716 GNDTKNIIRGLSPITNAGGVRRT-----EQNTITSDKSKAASIGDI 763
Db 566 KGDGAEL--ALDPS-PVYKNSGSGYATSKNYIGWSAIFDQGEAGAPAKVPIPNVGA 622
QY 764 LNTGFNLKNNNSNGVFS-TYNTWDFIDGNATTAKV-----TYDEFNOTSKVT--- 810
Db 623 QAPSAKGVNNGNATGYDLSHKAV---SGATGYKVQVFNKGKGFELDLIGNOTSWTTKG 678
QY 811 -----YDVNVDEKTIELTDGNGKINKIGVKTTLTTTNNANGKATNFS-----TTDND 858
Db 679 KIMPTSAEIRKAGYALHLKDGSGAELPIN-PGPTYKNAGDGAARVSKFIAYNKDGEA 737
QY 859 LVN-----AKDAENLNTLAKELHTTKGTADTALQTFVKKDGATDDETIIVGKDGTQ 912
Db 738 IASPAATPALDIPRKNVTGYLYTNK-----SSQT 769
QY 913 KTVNTLKLKGENGLTVATNKDGTVFQINTQSLKAGDS-----TTLNKDGLSKINPASNE 969
Db 770 G-YVNLWIEKVQNAKGYKVN---IYNGKEYQS-FDVGDAHWTTONKNLW-----PTSEE 819
QY 970 QIQVGADGVKFAVDK-----NSSTGIDGTSRITKQDQIGFT-----GANGSLD 1013
Db 820 ---IKAGSYKLHTDGGKELALDPSPVYNNANGNYKG-----KKNYSETLVAYDANGETI 871
QY 1014 TTKP-----HLTKDKLKVGE-----VEITNTGINAGCKKITNQSGLDITONSNDVATGGRV 1064
Db 872 PTAPPNPTFHEGAFFLTGEYWSLIIDPSQLNGA-----TGNVYNEEDLSIDGRG 923
QY 1065 YDLKTELESKINSAAKTAQNSLH-----EFSVADEQGNHFTVSNPYSSYDTSKTS 1114
Db 924 PGL-----GLSRYSNLSOSSDLHFGQGWADAETSIVSTDGAMKIDEATHTRETKA 977
QY 1115 DVITEAGENCI---TKYKNGVVRVIGDQT-----KGLTPKLVGNNNGKIGVIDSK 1164
Db 978 DG-TYOPPGVYLELTETADQFLLTK-BQTNAYFNKKGKGLQKVYDGHNNATVVTYNDK 1035
QY 1165 DQONTITGLSNTLANVNDPCAGHALS-QGLANDTDKTRAASIGDVLNAGFNQNGEAVD 1223
Db 1036 NQLTATDASGRKLTFTYDENGHVTSITGPKN--KKVTSYENDLLKKVTDGTVTSD 1093
QY 1224 FVSTYDVFIDGNATITAKVTYDD-----TSKTSKYVYDVNVNDKTIETVTS-- 1269
Db 1094 YDSEGRVLKQYSANSTEAPKVPFTEYOYSGHRLKATNAKETYVYSYDADKKTLLMTQPN 1153
QY 1270 -----DKLGVKTTTLTKTSANGNATKFSAADGDALVRASDIATHLN 1311
Db 1154 GRKVOGYNEAGNPQIVIDDAEGLKITNTK-----YEGNNVVEDVD----- 1195
QY 1312 TLAGDIQTAKGASQASSASYVDAGNKV-IYDSTDKKYQVNDKGOVDKKN-----E 1363
Db 1196 --PNDVGTGK---ATESYOY-DKDGNTSVKDAYGTETYEYKNNNDVTKKMDTEGNVTD 1248
QY 1364 VAKDKLVAQAOTPDGTLAQMNVKSYNK-----EQVNDANKKQGINEDNAFTKLE----- 1414
Db 1249 IAYDGLDAVSETDQS--GRSSAAVYDYKQYIOSSKLSASTNLLKDGSEFAOKSGWNL 1306
QY 1415 NAAKDTK-----TKNAAVTGDLNVAQOTPLTEAGDTGTAKKLGETLITKG 1462
Db 1307 TASKDRRKISVIADKSGVLGSKALEVLSQSISACTDHCYSSATQTVLEPNTTYTSLGK 1366
QY 1463 -QTDYTNK-----LTD-----NNTGVVAGTDFGTVKLAIDLNTLNSVNAAGTR 1503
Db 1367 IKTDLAKSRAYFNIDRLDKQRIOWIHNEYSALAGKNDWT---KQITFTTPANAKAV 1423
QY 1504 I-----DEKGISEVD---ANGQAKANTPVLSANGLDLGGRKISNTGAAYDD----- 1546
Db 1424 VYMEVDHKDKCKKAWFEVQLEKGEVSSYNPVQSSFTSATENMNVSASVDSBEGF 1483
QY 1547 NDVAVNFKQFNEVAKTVNNLNQNSGASLPEW--TDANGKPIGTGKPKOKA--IKGAD 1602
Db 1484 NDDVSLKAARTSASQAGSVTQOT-----VLGQSANDKPVYTLTGMSKASSVFTD 1535

QY 1603 GKYYHANANGVPVDKGPITDADKLANLAAGKPLDAGHOVVASLGGNSDAITLTNKS 1662
Db 1536 EXDYSLOAN-----VTYAD--GSTGIYNAKPPSGTQ-----EMNRAAVWIPKTK- 1577
QY 1663 TLPOIDTPTMNGNANAGAOASLPSLSAAQOSNAASV---KDVLENVGFNLQT-----NHNQ 1713
Db 1578 PINKVD-----ISILFQKSATGTWFWFDIRLIEGSLLTAKSYDYSNGY 1620
QY 1714 VDFVKAYDTVNFVNGTGADITSVRSAD-----GTMSNTVNV---TALAAT 1755
Db 1621 V--TKEEDELGTATSTDYDETCKTSETDAKGEKTYTYDQADQLTNMTLSNGTSLHSY 1678
QY 1756 DDDGNVL---IKAKDGKFKAD-DLM-----PNGSLKAGK-SASDAKTPGLSLVNP 1802
Db 1679 DREGNEVSKTIRAGADQYKFEYDVGKLVKTTDPLGNVLASEYDANSNLTKT-----ISP 1734
QY 1803 NAGKSTG---DAV---ALNNLSKAVFK-SKDGTMTTTSVSSDGSIOGK--DNSS-ITL 1851
Db 1735 NGNEVSLSYDGTDRVKSXSYNGTEKYIFETYDKNGNETSVNKEQNTTKRTFDKNRLTE 1794
QY 1852 SKDGLNVGKV-----ISNVGKGTDKTDAANVQOLNEVRNLLG----- 1889
Db 1795 LTDR---GQSQTWTPSDSKLKFESWIHGQKGTNQFTYKNKLDQMIEMKDSSTSYSPDY 1851
QY 1890 -----LGNAGNDNADQNVNIADIKK--DPNSG-----SSNRRTVVK--- 1924
Db 1852 DENGNVQTFITGGGTSFSYDERNLVSSLHIGDKNGGDILTESYEDANGNRRTINSSA 1911
QY 1925 AGTVL--CGKNNNDTEKATGGVQGVVDKDNANGDLSNVVWTKQKGSKKALLATYNA 1982
Db 1912 SKQVQYEGKLNQLVKETHEDGTVEITYDGFN---RKTVTVTIKDGSSTKTVNASENIM 1967
QY 1983 GQTNLYLTNPAEAIDRINEQGI-----REFHVND-----GNOQEPV- 2018
Db 1968 NQ-----LTKYNDESISYDKNCRNRTSDGKFTYTDADENLNAVTKKEDKPFAT 2016
QY 2019 ---QCRNGIDSSASGK-----HSVAIGFOAKADGEAAVAITGRQTOAGNQSIAGDN 2066
Db 2017 KYDEKGNRIQKTVNGKVTNYFYDGDLSNLVLYETDADNV-----TKSYTYGDS 2065
QY 2067 AQ-----ATGD-QSIAIGTGNVTVGKHSIGAIGDPSTVKA-----DNSYS 2104
Db 2066 QQLLSYTYENGKKYFHYNAHGDIIAISDSCTKVAKYOYDAWGNPTKTEASDEVKDNRYR 2125
QY 2105 VG-----NNNQFI-----DATQTDVFGVGNNTV-----TE 2130
Db 2126 YAGYQYDEETGLYLMARYEPRNGVFLSLDPPGSDGSDSLDQNGYAYGNNNPVWVNDPD 2185
QY 2131 SNSVALGSNSAISAGTHAGTQAKSDGTAGTTTTAGATGTVKGFAGQATVAGVSVGASGA 2190
Db 2186 GHWVWLNVNAGFAA--YDGYKAYKSG-----KGWKAAGAAAA----- 2220
QY 2191 ERRIQNVAAGEVSATSDAVNGSOLYKATOGIANATNELDRIHON-----ENK 2239
Db 2221 ---SNFGPKIKF-----CASPAYKFTKAVKITGHTRHGLNQSIGRNGGKGVNLR 2270
QY 2240 ANAGISSAMAMASMPQA---YIPGRSMVTGGIATHNGQGAVALGSKLSDNGQWFKING 2296
Db 2271 LNA-VRSPPKKVIQPNGATRYGVKKATV---VLNKRKGVITAYGSSR-AGKSKIHVFHTHG 2325
QY 2297 SAD 2299
Db 2326 KGN 2328

RESULT 13

PIP_LACLC

ID PIP_LACLC

AC PIP_LACLC

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE

STANDARD;

PRT; 1902 AA.

QY 458 GTDGLKFTDMSNTALEDTTTRITKDKIGFNSKACTVDENKPYLDKDKLVGNSTLNNGLT 517
 Db 57 ATDTAATTNQAIA-----TQLAAKGIDY-NLKNVQOQDIYVIVQMSAAPASENGTL- 110
 QY 518 VNNTIGGSNKIQGVAGDGKIFADVNWNVNAAKFGTTRITEEIGFADADGKV-----DKK 573
 Db 111 --RTDYSSTAEIQEINKVIAQASV-----KAAVEQVTOQTAG--ESYGVVNGFSTK 160
 QY 574 SPYLDKKQL-QVGGVKIT-----KDSGINA-----GDOKISNVKDATD 610
 Db 161 VRVVDIPKQLIAGVKTIVTLAKVYVPTDAKANSMAVQWNSYKYKGGTVVSVDSGI 220
 QY 611 DTDVATYKOLKOVQOQADGALOSFSIRDEKGOEFTISNLYSNCNTPTTETITFACENGI 670
 Db 221 DP---THKDMR-LUSDOKVLTQSDV--BKFTDTAKHGRYFNSKVPYGFN-----YADNND 270
 QY 671 SISNDIAKGVKVGIDIPINGLTTPKLTGSDKDKTQQLVIEQVAGSNDTKNIIRGLSP-- 728
 Db 271 TITDDT-----VDEQGHMVAGI-IGAN-----GTGDDPAKSVGVVAPEA 309
 QY 729 ---TLPSINAGVRRTEOGNTITSDEDSKSAASIGDIINTGFNLKNNNSNVGFVSTYNT 785
 Db 310 QLLAMKVFNTSDTSATGTSATLSAIEDSAKIGA--DVLNMSLG-SDSGNQTLDEPELAA 366
 QY 786 VDFIDGNATTAKVYDETQTSKVTYDVNVNDEKTIELTGDNKTKNIGIKVKTITLTTTAN 845
 Db 367 VQANESGTAAVISAGNSGTSATGKNDYVGLQ---DNEMVGPFG--TSRGATTVAS 421
 QY 846 GRATNESTTDNDALVNAKDIAENLNTLAKEIHTTKGTA-DTALQTEFKVKD---GATDDE 901
 Db 422 -----AENDVITQAVITDGTGLOLGPETIQLSSNDFGTSFDOK 461
 QY 902 TITVKGDTQN---GRTVN-TLKLKGENGLT-----VATNKDG 935
 Db 462 KFYVWDASGLSKGVADYADAKGKIAIKRGELTFADKOKYAQAAGAAGLIIVNNDG 521
 QY 936 TV-----TFGINTOSGLKAGDSTTLN-KDGLSIK----- 963
 Db 522 TATPVTSMALTTFTTFFGLSSVTGQKLDVMAAHPDDSLGVKIALTLVPNQRYTEDKMSD 581
 QY 964 ---NPASNEIQ--VGADGVKFAKVDKNSSTGIDGTSRITKDQIGFTGANGSL----- 1012
 Db 582 FTSYGPVSNLSPKPDITAGGNITWSTONNGYTNMSTSWASP-----FIAGSOALLKQAL 637
 QY 1013 -----DTPKPLHFKDLKVEVEITNTGINAGKKITINOSGDITQNSDVAIT- 1060
 Db 638 NNKNPNFYAYYKOLKGTALTDFLKTVMEN-TAOPIN-----DI--NYYNVIYS 682
 QY 1061 ---GGRVVDLKTLES-KINSAKTAQN-----SLHEFSVADBO-----GNHFTVSNPY 1105
 Db 683 PRQAGLVVVKAAIDALEKNPSTVVAENGYPAVELKDFSTDKTKFLTFTNRTHELTY 742
 QY 1106 SSYDTSKTSVITFAGE--NGI--TTKVNKVVVRVIGDQT----- 1141
 Db 743 QMDSNTDTNAVYTSATDPNSGVLYDKKIDGAAIKAGSNITVPAGKTAQIEFTLSLPKSED 802
 QY 1142 -----KGLTPPKLTV-----GN-NNGKGVIVDSKDG--QNTITGLSNTLANVT 1181
 Db 803 QOQFVEGFLNFKGSDGSRNLNPMYGFWDNDGK--IVDSLNGITYSPAGNFGTVPLLT 860
 QY 1182 NDGAGHALSQGLANDPDKTRAASIGDVLNAGFNLOQNGCAVDFVSYDIVDFIDGNATTA 1241
 Db 861 NKWTGYGGMWVTDADGNQTV-----DQQAIAFSS-----DKNA--- 895
 QY 1242 KVTYDDTSKTSKVYDVNVNDNKITEVTSDKLGVKVTTLTKTSANGNATKFSAADGDAL- 1300
 Db 896 --LYNDIS--MKYLLRLNSNVQVDILDCQ--GNKVTLLS-SSTRKKKYYNAHSQQYIY 948
 QY 1301 --VKASDIATHLNTLAGDIOTAKGASQASSASYSVDADGNKVIYD---STDKKIYQV--- 1352
 Db 949 YHAPAWD-GTYVDORDGNIKTADGDSYTYRISGVPEGGKQGVDFVPPFKLDSKAPTVRHV 1007
 QY 1353 -----NDKGQVDKKNKEVAKDKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANKKOGINE 1405

RESULT 15
 CNA_STAAU
 ID CNA_STAAU
 AC Q53654;
 DT 15-DEC-1998 (Rel. 37, Created)

STANDARD;
 PRT; 1183 AA.

Db 1008 ALSAKTENCKTQTYLLTAE-AKDDL-----SGLDATKSVKTAINEVTLNDATFTDAGTT 1059
 QY 1406 DNAFIKGLENAADTKTKNAAVTVGDLNVAQOTPLT-----FAGDTGTAKKLCGET---LT 1458
 Db 1060 ADGYTK-IETPLSDEQ-----AQALGNGDNSAEIYLTDNASNATDQDASVOKPOSTSFDLI 1114
 QY 1459 IKGGQDTPNKLTNNNIGVAGTDG-----FTVKLAKDLTNLSNVNAGGTRIDEKG--- 1508
 Db 1115 VNGGGI-PDKISSTTTGYEANTOGGGTYTFESGYPAADV-----GTYTDAQKKH 1163
 QY 1509 ---ISFVDANGQAKANTPVLSANGLDLGCKRISNIGAAVDDNDNAVNFQKQNE-----VAKT 1561
 Db 1164 DLNNTTYDAATNSPTASMPVTNAD-----YAAQVLDYADKAHTQLKLFHFDTKVRLTAPT 1216
 QY 1562 VNNL--NNOSNCGASLPFVVVT---DANGKPIGTGDKPQKAIKAGDGKYYHANANGVPV- 1615
 Db 1217 FTDLKENNGSDQTSSEATIKVTGVSADTKTVNVGD-----TVAALDAQ--HHSVDVVPVN 1269
 QY 1616 -----DKDG-----KPIT---DADKLANIAAHGKPLDAGHQVVVASLGGNSDAI 1655
 Db 1270 YGDNITIKVTATDEGNTTTEQKTTISYDPDKLNSVT-----FDQG-----VTFGAN--- 1317
 QY 1656 TLTNIKSTLPOIDTPNTGNANAGQASLPSLSAAQOSNAASVKDVLNVGNL----- 1707
 Db 1318 ---EFNATSAKFYDPKGTGATITGKVKHPTTTLQVDPKGQPIKIDDLTFSFTLDGLTGOK 1374
 QY 1708 -----QTNHNOVDFVKAYDTVNFVNGTGADITSV-RSADGTMSNITVTALAAATDDDG 1759
 Db 1375 PFCVVVGDITQNTK---FQEAFTFILDVAVPTLSLSDSDTADPVTNNPFOITGT--- 1426
 QY 1760 NVLIKAKDKFYKADLMPNGSLKAGKSASDAKTPTGSLVNPNAKG---STGDAVAL- 1815
 Db 1427 -----ATDNAQYLS--LSINGSSVASQYVD-----ININSGKPGHMAIDQPVKLL 1469
 QY 1816 ---NNLSKAVFKSKDGTITTTTSSDGIISQCKDN-----SSITLSKQGLNV 1858
 Db 1470 EGNVLTVAVTQSEDNNTTKNIT---VYEPKKTLAAPTVPSTTPEAKTIVTLTANSAAT 1526
 QY 1859 GGRV--ISNVGKGTCTDAAVQ--QLNEVRNLLGLGNAGNDNAGNOVNTADIKDP--- 1912
 Db 1527 GETVQVYADGGKTYQDVPAAGVTVTANGTFKFKSTDLGYNESPAVDYV-VTNIKADDDPAQ 1585
 QY 1913 -NSGSSNRTVTKAGTVLGGKGNND-----TEKATG--GVQGVGVDKDGNGNDL 1959
 Db 1586 LQAAKQELNLNLSAKTSLASGKYDDATTTALAAATKQACTALDQTNASVDSLGTGANRDL 1645
 QY 1960 SNVVVTKQDKGSKALLATYNAAGOTNYLTNNPAAEIDRINEOGIRFFHVNDGNOBPVQ 2019
 Db 1646 QT-----AINQLAAKLPADKKTSL--NQLQSVKAALETDL-----GNQ----- 1682
 QY 2020 GRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQAATGDOSTAGT 2079
 Db 1683 ---TDSSTGK-----TFTAALDDLVA-----QAQACTQT---DQOLQATLAKVLDAVL 1724
 QY 2080 GNVVTKHSGAIGDPSTVKADNSYSVGNNNQFIDATOTDVFVGNNITVTYESVALGSN 2139
 Db 1725 AKLAEG-----IKAAATPAEVGNK---DAATGKTWYADIADTLTSGQASADAD 1770
 QY 2140 -----SAISACTHAGTQAKKSDGTAGTTTTAGATGT---VKGFAGOTAYGAVS 2184
 Db 1771 KLAHQALOSLTKTVAAGAAVEAAKTVGKGDTGTCTSKGGQGTGTPPTGDIKDGDEGS 1830
 QY 2185 VGASGAERRIQNVAAAGEVSATST---DAVNGSOLYKATOGIANATFELDRH 2232
 Db 1831 QPSSGG-----NIPTPATTTSTDDTDRNCQLISGKALPKTGETTER 1876

THIS PAGE BLANK (USPTO)